

OM protein - protein search, using sw model

Run on: February 27, 2004, 06:40:42 ; Search time 46.5363 Seconds  
(without alignments)  
3958.655 Million cell updates/sec

Title: US-09-989-981A-2  
Perfect score: 3369  
Sequence: 1 MGELPFLSPEGARGPHINRG.....PALVILGIVIFKVRDYLI SR 652

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	3369	100.0	652	5	AAE13289	Aae13289 Mouse sit
2	3369	100.0	652	6	AAE31702	Aae31702 Mouse ABC
3	3367	99.9	652	5	AAE13308	Aae13308 Mouse sit
4	3363	99.8	652	5	AAU96985	Aau96985 Mouse ABC
5	3363	99.8	652	5	AAE13309	Aae13309 Mouse sit
6	3150	93.5	652	5	AAU96986	Aau96986 Rat ABCG5
7	2744.5	81.5	651	5	AAU96984	Aau96984 Human ABC
8	2744.5	81.5	651	5	AAE13290	Aae13290 Human sit
9	2744.5	81.5	651	6	AAE31704	Aae31704 Human ABC

10	2741.5	81.4	651	5	AAU96992	Aau96992	Human	ABC
11	2739.5	81.3	651	5	AAU96990	Aau96990	Human	ABC
12	2739.5	81.3	651	5	AAU96989	Aau96989	Human	ABC
13	2737.5	81.3	651	5	AAU96993	Aau96993	Human	ABC
14	2722.5	80.8	649	5	ABP52128	Abp52128	Homo sapi	
15	1618.5	48.0	408	5	AAU96991	Aau96991	Human	ABC
16	1536.5	45.6	340	5	AAU96987	Aau96987	Hamster A	
17	1177	34.9	256	7	ADB64641	Adb64641	Human pro	
18	914.5	27.1	243	5	AAU96988	Aau96988	Human	ABC
19	707.5	21.0	672	6	AAE31703	Aae31703	Mouse	ABC
20	693.5	20.6	655	5	AAU80029	Aau80029	Human	ABC
21	693.5	20.6	663	2	AAY15221	Aay15221	Breast Ca	
22	691.5	20.5	655	4	AAB60104	Aab60104	Human tra	
23	691.5	20.5	655	5	AAO14781	Aao14781	Human	BCR
24	691.5	20.5	655	5	AAU80028	Aau80028	Human	ABC
25	691.5	20.5	655	6	ABR58077	Abr58077	Human	ABC
26	691.5	20.5	655	6	ADA10917	Ada10917	Human	cDN
27	691.5	20.5	655	7	ADC54182	Adc54182	Human	bre
28	690.5	20.5	665	5	AAO14782	Aao14782	Human	BCR
29	689.5	20.5	655	3	AAY95365	Aay95365	ATP-bindi	
30	689.5	20.5	655	4	AAU04348	Aau04348	Human	BCR
31	689.5	20.5	655	5	ABP52127	Abp52127	Homo sapi	
32	689.5	20.5	655	5	ABB07270	Abb07270	Human	BCR
33	689.5	20.5	655	6	ABU63376	Abu63376	Human	mit
34	688.5	20.4	673	5	ABP52129	Abp52129	Homo sapi	
35	688.5	20.4	673	6	AAE31705	Aae31705	Human	ABC
36	687.5	20.4	665	5	AAO14783	Aao14783	Human	BCR
37	683.5	20.3	655	5	ABB07273	Abb07273	Human	BCR
38	677	20.1	657	5	ABB07272	Abb07272	Murine	BC
39	640	19.0	144	3	AAB41856	Aab41856	Human	ORF
40	595.5	17.7	602	4	ABB65432	Abb65432	Drosophil	
41	592.5	17.6	674	5	ABP52126	Abp52126	Homo sapi	
42	589	17.5	687	4	ABB59384	Abb59384	Drosophil	
43	580	17.2	638	5	ABB98349	Abb98349	Human	ABC
44	576.5	17.1	646	5	AAE28964	Aae28964	Human	ABC
45	576.5	17.1	646	5	ABB98348	Abb98348	Human	ABC

# ALIGNMENTS

RESULT 1

AAE13289

ID AAE13289 standard; protein; 652 AA.

XX

AC AAE13289;

XX

DT 12-FEB-2002 (first entry)

XX

DE Mouse sitosterolaemia susceptibility gene (SSG) protein.

XX

KW Mouse; sitosterolaemia susceptibility gene; SSG; atherosclerosis;

KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia;

KW gall stone; coronary heart disease; cardiovascular disease; arthritis;

KW xanthoma; haemolytic anaemia; transgenic animal; chromosome 17; therapy.

XX

OS Mus sp.

XX  
 PN WO200179272-A2.  
 XX  
 PD 25-OCT-2001.  
 XX  
 PF 18-APR-2001; 2001WO-US012758.  
 XX  
 PR 18-APR-2000; 2000US-0198465P.  
 PR 15-MAY-2000; 2000US-0204234P.  
 XX  
 PA (TULA-) TULARIK INC.  
 XX  
 PI Tian H, Schultz J, Shan B;  
 XX  
 DR WPI; 2002-017598/02.  
 DR N-PSDB; AAD22008.  
 XX  
 PT Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,  
 PT useful for screening a compound that increases the level of expression or  
 PT activity of SSG polypeptide for treating sterol-related disorder.  
 XX  
 PS Claim 19; Fig 7; 105pp; English.  
 XX  
 CC The invention relates to an isolated Sitosterolaemia Susceptibility Gene  
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)  
 CC binding cassette (ABC) family cholesterol transporter. SSG is useful for  
 CC identifying a compound useful in the treatment or prevention of a sterol-  
 CC related disorder, including sitosterolaemia, hyperlipidaemia,  
 CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or  
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-  
 CC associated diseases or conditions including coronary heart disease and  
 CC other cardiovascular diseases, and sitosterolaemia-associated condition  
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG  
 CC expression cassette is useful in the production of transgenic non-human  
 CC animals. SSG genes and their homologues are useful as tools for a number  
 CC of applications including diagnosing sitosterolaemia and other  
 CC cardiovascular disorders, for forensics and paternity determinations, and  
 CC for treating any of a large number of SSG associated diseases. The  
 CC present sequence is mouse SSG protein. Mouse SSG is located on chromosome  
 CC 17  
 XX  
 SQ Sequence 652 AA;

Query Match 100.0%; Score 3369; DB 5; Length 652;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-313;  
 Matches 652; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1 MGELPFLSPEGARGPHINRGSLSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS 60
          |||
Db      1 MGELPFLSPEGARGPHINRGSLSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS 60

Qy     61 CQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVC 120
          |||
Db     61 CQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVC 120

Qy    121 LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNKKVEAVMTLSLSH 180
          |||

```

Db	121	LRRDQFQDCFSYVLQSDVFLSSSLTVRETTRYTAMLALCRSSADFYNNKKVEAVMTLSLSH	180
Qy	181	VADQMIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA	240
Db	181	VADQMIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA	240
Qy	241	RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPF	300
Db	241	RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPF	300
Qy	301	FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPFKTK	360
Db	301	FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPFKTK	360
Qy	361	DPPGMFGKLGVLRRVTRNLNRNQAVIMRLVQNLIMGLFLIFYLLRVQNNLTGKAVQDR	420
Db	361	DPPGMFGKLGVLRRVTRNLNRNQAVIMRLVQNLIMGLFLIFYLLRVQNNLTGKAVQDR	420
Qy	421	VGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLPFVSIAT	480
Db	421	VGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLPFVSIAT	480
Qy	481	VIFSSVCYWTGLYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPINVSIVALLSI	540
Db	481	VIFSSVCYWTGLYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPINVSIVALLSI	540
Qy	541	SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCILVVNEFYGLNFTCGGSNTSMLNHMP	600
Db	541	SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCILVVNEFYGLNFTCGGSNTSMLNHMP	600
Qy	601	CAITQGVQFIEKTCPGATSRFTANFLILYGFIPALVILGIVIFKVRDYLISR	652
Db	601	CAITQGVQFIEKTCPGATSRFTANFLILYGFIPALVILGIVIFKVRDYLISR	652

RESULT 2

AAE31702

ID AAE31702 standard; protein; 652 AA.

XX

AC AAE31702;

XX

DT 24-MAR-2003 (first entry)

XX

DE Mouse ABCG5 protein.

XX

KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;  
KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;  
KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;  
KW mouse; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;  
KW ABCG5.

XX

OS Mus sp.

XX

PN WO200281691-A2.

XX

PD 17-OCT-2002.

XX

PF 20-NOV-2001; 2001WO-US043823.

XX

PR 20-NOV-2000; 2000US-0252235P.

PR 28-NOV-2000; 2000US-0253645P.

XX

PA (TULA-) TULARIK INC.

PA (TEXA ) UNIV TEXAS SYSTEM.

XX

PI Hobbs HH, Shan B, Barnes R, Tian H;

XX

DR WPI; 2003-058548/05.

DR N-PSDB; AAD48880.

XX

PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-related disorders e.g. sitosterolemia, hypercholesterolemia, hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or nutritional deficiencies.

XX

PS Claim 28; Page 74; 94pp; English.

XX

CC The invention relates to ATP-binding cassette (ABC) family cholesterol transporter, ABCG8 polypeptides and polynucleotides. The invention also provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known as sitosterolaemia susceptibility gene (SSG). Sequences of the invention are useful for treating or preventing sterol-related disorders such as sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis and nutritional deficiencies. They are also useful in gene therapy. The present sequence is mouse ABCG5 protein

XX

SQ Sequence 652 AA;

Query Match 100.0%; Score 3369; DB 6; Length 652;  
Best Local Similarity 100.0%; Pred. No. 1.1e-313;  
Matches 652; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGELPFLSPEGARGPHINRGSLSSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS 60  
|  
Db 1 MGELPFLSPEGARGPHINRGSLSSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS 60

Qy 61 CQQKQWDRQILKDVSLYIESGQIMCILGSSSGSKTTLLDAISGRLRRTGTLEGEVFNVC 120  
|  
Db 61 CQQKQWDRQILKDVSLYIESGQIMCILGSSSGSKTTLLDAISGRLRRTGTLEGEVFNVC 120

Qy 121 LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKVEAVMTLSLSH 180  
|  
Db 121 LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKVEAVMTLSLSH 180

Qy 181 VADQMIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240  
|  
Db 181 VADQMIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240

Qy 241 RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCCGYPCPEHSNPF 300  
|  
Db 241 RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCCGYPCPEHSNPF 300

Qy 301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPFKTK 360  
|

```

Db      301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPVFFKTK 360
Qy      361 DPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLMGLFLIFYLLRVQNNLTGKAVQDR 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 DPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLMGLFLIFYLLRVQNNLTGKAVQDR 420
Qy      421 VGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLPFSVIAT 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 VGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLPFSVIAT 480
Qy      481 VIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 VIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI 540
Qy      541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPM 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPM 600
Qy      601 CAITQGVQFIEKTCPGATSREFTANFLILYGFIPALVILGIVIFKVRDYLSR 652
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 CAITQGVQFIEKTCPGATSREFTANFLILYGFIPALVILGIVIFKVRDYLSR 652

```

RESULT 3

AAE13308

ID AAE13308 standard; protein; 652 AA.

XX

AC AAE13308;

XX

DT 12-FEB-2002 (first entry)

XX

DE Mouse sitosterolaemia susceptibility gene (SSG) protein variant #1.

XX

KW Mouse; sitosterolaemia susceptibility gene; SSG; atherosclerosis; mutein;

KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; mutant;

KW gall stone; coronary heart disease; cardiovascular disease; arthritis;

KW xanthoma; haemolytic anaemia; transgenic animal; therapy; variant.

XX

OS Mus sp.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Misc-difference 17

FT /note= "Wild type Ile substituted with Leu"

XX

PN WO200179272-A2.

XX

PD 25-OCT-2001.

XX

PF 18-APR-2001; 2001WO-US012758.

XX

PR 18-APR-2000; 2000US-0198465P.

PR 15-MAY-2000; 2000US-0204234P.

XX

PA (TULA-) TULARIK INC.

XX

PI Tian H, Schultz J, Shan B;

XX  
DR WPI; 2002-017598/02.  
XX  
PT Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,  
PT useful for screening a compound that increases the level of expression or  
PT activity of SSG polypeptide for treating sterol-related disorder.  
XX  
PS Disclosure; Page; 105pp; English.  
XX  
CC The invention relates to an isolated Sitosterolaemia Susceptibility Gene  
CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)  
CC binding cassette (ABC) family cholesterol transporter. SSG is useful for  
CC identifying a compound useful in the treatment or prevention of a sterol-  
CC related disorder, including sitosterolaemia, hyperlipidaemia,  
CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or  
CC nutritional deficiencies. SSG is also useful for treating cholesterol-  
CC associated diseases or conditions including coronary heart disease and  
CC other cardiovascular diseases, and sitosterolaemia-associated condition  
CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG  
CC expression cassette is useful in the production of transgenic non-human  
CC animals. SSG genes and their homologues are useful as tools for a number  
CC of applications including diagnosing sitosterolaemia and other  
CC cardiovascular disorders, for forensics and paternity determinations, and  
CC for treating any of a large number of SSG associated diseases. The  
CC present sequence is mouse SSG protein variant obtained by replacing Ile17  
CC with Leu. Note: The present sequence is not shown in the specification  
CC but is derived from mouse SSG protein referred as SEQ ID NO: 1 (AAE13289)  
CC and shown in figure 7 of the specification  
XX  
SQ Sequence 652 AA;

Query Match 99.9%; Score 3367; DB 5; Length 652;  
Best Local Similarity 99.8%; Pred. No. 1.7e-313;  
Matches 651; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MGELPFLSPEGARGPHINRGSLSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS	60
		:	
Db	1	MGELPFLSPEGARGPHLNRSLSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS	60
Qy	61	CQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVC	120
Db	61	CQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVC	120
Qy	121	LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTLSLSH	180
Db	121	LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTLSLSH	180
Qy	181	VADQMIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA	240
Db	181	VADQMIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA	240
Qy	241	RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCCGYPCPEHSNPF	300
Db	241	RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCCGYPCPEHSNPF	300
Qy	301	FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPVFPKTK	360

Db 301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPFVKTK 360  
 Qy 361 DPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLMGLFLIFYLLRVQNNTLKGAQQDR 420  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 361 DPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLMGLFLIFYLLRVQNNTLKGAQQDR 420  
 Qy 421 VGLLYQLVGATPYTGMLNAVNLFPMRAVSDQESQDGLYHKWQMLLAYVLHVLPFSVIAT 480  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 421 VGLLYQLVGATPYTGMLNAVNLFPMRAVSDQESQDGLYHKWQMLLAYVLHVLPFSVIAT 480  
 Qy 481 VIFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI 540  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 481 VIFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI 540  
 Qy 541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHMP 600  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHMP 600  
 Qy 601 CAITQGVQFIEKTCPGATSRETNANFLILYGFIPALVILGIVIFKVRDYLISS 652  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 601 CAITQGVQFIEKTCPGATSRETNANFLILYGFIPALVILGIVIFKVRDYLISS 652

RESULT 4

AAU96985

ID AAU96985 standard; protein; 652 AA.

XX

AC AAU96985;

XX

DT 30-JUL-2002 (first entry)

XX

DE Mouse ABCG5 protein.

XX

KW Mouse; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;  
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease.

XX

OS Mus sp.

XX

FH Key Location/Qualifiers

FT Misc-difference 638. .652

FT /note= "Encoded by CTAG"

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.



DR N-PSDB; ABK51684.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic  
PT acid encoding the polypeptide, useful for treating sitosterolemia,  
PT arteriosclerosis and heart diseases.

XX

PS Example 3; Page 42; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette  
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a  
CC predisposition for developing sitosterolemia, arteriosclerosis or heart  
CC disease. The molecules of the invention are also useful for identifying a  
CC compound which alters ABCG5 activity level comprising contacting a cell  
CC culture or mammal which have ABCG5 polypeptide with a compound and  
CC measuring ABCG5 biological activity in the cell culture or in mammal,  
CC where an increase or decrease in ABCG5 biological activity compared to  
CC ABCG5 biological activity in a control cell culture or mammal not  
CC contacted with the compound, identifies a compound that increases or  
CC decreases ABCG5 activity respectively. The cell culture or mammal  
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The  
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the  
CC polypeptide in a cell culture or mammal is also compared with that of a  
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.  
CC Stimulation of ABCG5 activity is useful for treating or preventing  
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's  
CC disease. The method of the invention is useful for increasing cholesterol  
CC excretion and/or decreasing cholesterol adsorption. The present amino  
CC acid sequence represents the mouse ABCG5 protein of the invention

XX

SQ Sequence 652 AA;

Query Match 99.8%; Score 3363; DB 5; Length 652;  
Best Local Similarity 99.8%; Pred. No. 4e-313;  
Matches 651; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MGELPFLSPEGARGPHINRGSLSSLEQGSVTGT	TEARHSLGVLHVSYSVSNRVGPWWNIKS	60
Db	1	MGELPFLSPEGARGPHINRGSLSSLEQGSVTGT	TEARHSLGVLHVSYSVSNRVGPWWNIKS	60
Qy	61	CQQKWDRQILKDVS	LYIESGQIMCILGSSSGSKTLLDAISGRLRRTGTLEGEV	120
Db	61	CQQKWDRQILKDVS	LYIESGQIMCILGSSSGSKTLLDAISGRLRRTGTLEGEV	120
Qy	121	LRRDQFQDCFSYVLQSDVFLSSLT	TVRETLRYTAMLALCRSSADFYNNKKVEAVMT	180
Db	121	LRRDQFQDCFSYVLQSDVFLSSLT	TVRETLRYTAMLALCRSSADFYNNKKVEAVMT	180
Qy	181	VADQMIGSYNFGGISSGERRRV	SIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA	240
Db	181	VADQMIGSYNFGGISSGERRRV	SIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA	240
Qy	241	RRDRIVIVTIHQPRSELFQHF	DKIAILTYGELVFCGTPEEMLGFFN	300
Db	241	RRDRIVIVTIHQPRSELFQHF	DKIAILTYGELVFCGTPEEMLGFFN	300
Qy	301	FYMDLTSVDTQSREREIETYKRVQ	MLECAFKESDIYHKILENIERARYLKTLPMPVFKTK	360

```

Db      301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPVPFKTK 360
Qy      361 DPPGMFGKLGVLRLRRVTRNLMRNKQAVIMRLVQNLMGLFLIFYLLRVQNNTLKGAQDR 420
        |||
Db      361 DPPGMFGKLGVLRLRRVTRNLMRNKQAVIMRLVQNLMGLFLIFYLLRVQNNTLKGAQDR 420
Qy      421 VGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYVLHVLPPFSVIAT 480
        |||
Db      421 VGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYVLHVLPPFSVIAT 480
Qy      481 VIFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI 540
        |||
Db      481 VIFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI 540
Qy      541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVNEFYGLNFTCGGSNTSMLNHMP 600
        |||
Db      541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVNEFYGLNFTCGGSNTSMLNHMP 600
Qy      601 CAITQGVQFIEKTCPGATSRTANFLILYGFIPALVILGIVIFKVRDYILSR 652
        |||
Db      601 CAITQGVQFIEKTCPGATSRTANFLILYGFIPALVILGIVIFKVRDYILSR 652

```

RESULT 5

AAE13309

ID AAE13309 standard; protein; 652 AA.

XX

AC AAE13309;

XX

DT 12-FEB-2002 (first entry)

XX

DE Mouse sitosterolaemia susceptibility gene (SSG) protein variant #2.

XX

KW Mouse; sitosterolaemia susceptibility gene; SSG; atherosclerosis; mutein;

KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; mutant;

KW gall stone; coronary heart disease; cardiovascular disease; arthritis;

KW xanthoma; haemolytic anaemia; transgenic animal; therapy; variant.

XX

OS Mus sp.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Misc-difference 28

FT /note= "Wild type Gly substituted with Ala"

XX

PN WO200179272-A2.

XX

PD 25-OCT-2001.

XX

PF 18-APR-2001; 2001WO-US012758.

XX

PR 18-APR-2000; 2000US-0198465P.

PR 15-MAY-2000; 2000US-0204234P.

XX

PA (TULA-) TULARIK INC.

XX

PI Tian H, Schultz J, Shan B;



Db 301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPVFFKTK 360  
 Qy 361 DPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLMGLFLIFYLLRVQNNTLKGAQQDR 420  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 361 DPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLMGLFLIFYLLRVQNNTLKGAQQDR 420  
 Qy 421 VGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLPPFSVIAT 480  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 421 VGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLPPFSVIAT 480  
 Qy 481 VIFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI 540  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 481 VIFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI 540  
 Qy 541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPM 600  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPM 600  
 Qy 601 CAITQGVQFIEKTCPGATSRETNANFLILYGFIPALVILGIVIFKVRDYILSR 652  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 601 CAITQGVQFIEKTCPGATSRETNANFLILYGFIPALVILGIVIFKVRDYILSR 652

RESULT 6

AAU96986

ID AAU96986 standard; protein; 652 AA.

XX

AC AAU96986;

XX

DT 07-AUG-2003 (revised)

DT 30-JUL-2002 (first entry)

XX

DE Rat ABCG5 protein.

XX

KW Rat; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;  
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease.

XX

OS Rattus sp.

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

DR N-PSDB; ABK51686.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic







Db 300 FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSAICHKTLKNIERMKHLKTLPMVPFKTK 359  
 Qy 361 DPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNTLKGAVQDR 420  
 | ||:| |||||:|||| ||| ||:|||||:|::|||:| ||||:||||  
 Db 360 DSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLVRVSNVLKGAIQDR 419  
 Qy 421 VGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLFPFSVIAT 480  
 ||||| |||||:||||| ||||:|||| |||||:||  
 Db 420 VGLLYQFVGATPYTGMLNAVNLFVLRVAVSDQESQDGLYQKWQMLLAYALHVLFPFSVVAT 479  
 Qy 481 VIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI 540  
 :|||||:|||||:|||||:|||||:|||||:|||||  
 Db 480 MIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSVVALLSI 539  
 Qy 541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHMP 600  
 :||:||||:||||| ||: ||||| ||||| ||||| |||: ||  
 Db 540 AGVLVSGGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPM 599  
 Qy 601 CAITQGVQFIEKTCPGATSRTANFLILYGFIPALVILGIVIFKVRDYILISR 652  
 || |||:||||| ||||| ||||| |||||:||:||:||||  
 Db 600 CAFTQGIQFIEKTCPGATSRTMNFLILYSFIPALVILGIVVFKIRDHLISR 651

RESULT 8

AAE13290

ID AAE13290 standard; protein; 651 AA.

XX

AC AAE13290;

XX

DT 12-FEB-2002 (first entry)

XX

DE Human sitosterolaemia susceptibility gene (SSG) protein.

XX

KW Human; sitosterolaemia susceptibility gene; SSG; atherosclerosis;  
 KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; therapy;  
 KW gall stone; coronary heart disease; cardiovascular disease; arthritis;  
 KW xanthoma; haemolytic anaemia; transgenic animal; chromosome 2p21.

XX

OS Homo sapiens.

XX

PN WO200179272-A2.

XX

PD 25-OCT-2001.

XX

PF 18-APR-2001; 2001WO-US012758.

XX

PR 18-APR-2000; 2000US-0198465P.

PR 15-MAY-2000; 2000US-0204234P.

XX

PA (TULA-) TULARIK INC.

XX

PI Tian H, Schultz J, Shan B;

XX

DR WPI; 2002-017598/02.

DR N-PSDB; AAD22009.

XX

PT Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,



PT useful for screening a compound that increases the level of expression or  
PT activity of SSG polypeptide for treating sterol-related disorder.

XX

PS Claim 19; Fig 8; 105pp; English.

XX

CC The invention relates to an isolated Sitosterolaemia Susceptibility Gene  
CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)  
CC binding cassette (ABC) family cholesterol transporter. SSG is useful for  
CC identifying a compound useful in the treatment or prevention of a sterol-  
CC related disorder, including sitosterolaemia, hyperlipidaemia,  
CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or  
CC nutritional deficiencies. SSG is also useful for treating cholesterol-  
CC associated diseases or conditions including coronary heart disease and  
CC other cardiovascular diseases, and sitosterolaemia-associated condition  
CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG  
CC expression cassette is useful in the production of transgenic non-human  
CC animals. SSG genes and their homologues are useful as tools for a number  
CC of applications including diagnosing sitosterolaemia and other  
CC cardiovascular disorders, for forensics and paternity determinations, and  
CC for treating any of a large number of SSG associated diseases. The  
CC present sequence is human SSG protein. Human SSG is located on chromosome  
CC 2p21

XX

SQ Sequence 651 AA;

Query Match 81.5%; Score 2744.5; DB 5; Length 651;  
Best Local Similarity 80.2%; Pred. No. 8.6e-254;  
Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;

Qy 1 MGELPFLSPEGARGPHINRGSLSLEQGSVTGTPEARHSLGVLHVSYSVSNRVGPWWNIKS 60  
|:| |:| |:| :||| ||| | | |||:| |||:| |  
Db 1 MGDLSLTPGGSMGLQVNRSQSSLEGAPATAPEP-HSLGILHASYSVSHRVRPWWDITS 59

Qy 61 CQQKWDQRILKDVSLYIESGQIMCILGSSSGSGKTTLLDAISGRLRRTGTLEGEVFNVCCE 120  
|:|:| |||||:|||||:|||||:|||| | || |||:|  
Db 60 CRQQWTRQILKDVSLYVESGQIMCILGSSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRA 119

Qy 121 LRRDQFQDCFSYVLQSDVFLSSLTVRETLYRTAMLALCRSSADFYNKKVEAVMTLSLSH 180  
|||:||||| ||||| |||:|: | : : ||||| |||||  
Db 120 LRREQFQDCFSYVLQSDTLLSSLTVRETLYHTALLAIRGNPGSFQKKVEAVMAELSLSH 179

Qy 181 VADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240  
|||:|:|:|: ||||:|||||:|||||: |||||: |||||: |||  
Db 180 VADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMMLFDEPTTGLDCMTANQIVVLLVEIA 239

Qy 241 RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCCGYPCPEHSNPF 300  
||:|:|:|:| |||||:|:|:|: ||| |||:|:|:|: |||||  
Db 240 RRNRIVVLTIHQPRSELFQLFQDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPF 299

Qy 301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPVFKTK 360  
|||||:||||| ||||:| |:|:| | | |:| | :|:|:|:|  
Db 300 FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMPVFKTK 359

Qy 361 DPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNTLKGAVQDR 420  
| |:| |||||:| ||| |:|:|:|:|:|:|:| |:|:|  
Db 360 DSPGVFSLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDR 419

Qy	421	VGLLYQLVGATPYTGMLNAVNLFPM LRAVSDQESQDGLYHKWQMLLAYVLHVL PFSVIAT	480
		:                     :                     :	
Db	420	VGLLYQFVGATPYTGMLNAVNLF PVLRAVSDQESQDGLYQKWQMMLAYALHVL PFSVVAT	479
Qy	481	VIFSSVCYWTLGLYPEVARFGYFSAALLAPHLIGEF LTLVLLGIVQNPNIVNSIVALLSI	540
		:                 :                                   :	
Db	480	MIFSSVCYWTLGLHPEVARFGYFSAALLAPHLIGEF LTLVLLGIVQNPNIVNSSVALLSI	539
Qy	541	SGLLIGSGFIRNIQEMPIPLKILGYFT FQKYCCEILVVNEFYGLNFTCGGSNTSMLNH PM	600
		:   :   :     :               :                           : :	
Db	540	AGVLVSGSFLRNIQEMPIPFKIISYFT FQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPM	599
Qy	601	CAITQG VQFIEKTCPGATS RFTANFLILYGFI PALVILGIVIFKVRDY LISR	652
		:                             :   :   :	
Db	600	CAFTOGIQFIEKTCPGATS RFTMNFILYSFI PALVILGIVVF KIRDHLISR	651

## RESULT 9

AAE31704

ID AAE31704 standard; protein; 651 AA.

XX

AC AAE31704;

XX

DT 24-MAR-2003 (first entry)

XX

DE Human ABCG5 protein.

XX

KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;

KW sitosteralaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;

KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;

KW human; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;

KW ABCG5.

XX

OS Homo sapiens.

XX

PN WO200281691-A2.

XX

PD 17-OCT-2002.

XX

PF 20-NOV-2001; 2001WO-US043823.

XX

PR 20-NOV-2000; 2000US-0252235P.

PR 28-NOV-2000; 2000US-0253645P.

XX

PA (TULA-) TULARIK INC.

PA (TEXA ) UNIV TEXAS SYSTEM. .

XX

PI Hobbs HH, Shan B, Barnes R, Tian H;

XX

DR WPI; 2003-058548/05.

DR N-PSDB; AAD48882.

XX

PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-

PT related disorders e.g. sitosterolemia, hypercholesterolemia,

PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or

PT nutritional deficiencies.

XX

PS Claim 28; Page 78-79; 94pp; English.

XX

CC The invention relates to ATP-binding cassette (ABC) family cholesterol  
CC transporter, ABCG8 polypeptides and polynucleotides. The invention also  
CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known  
CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention  
CC are useful for treating or preventing sterol-related disorders such as  
CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL  
CC deficiency, atherosclerosis and nutritional deficiencies. They are also  
CC useful in gene therapy. The present sequence is human ABCG5 protein

XX

SQ Sequence 651 AA;

Query Match 81.5%; Score 2744.5; DB 6; Length 651;  
Best Local Similarity 80.2%; Pred. No. 8.6e-254;  
Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;

Qy 1 MGELPFLSPEGARGPHINRGSLSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS 60  
||:| |:| |:| :||| ||| | | |||:| ||||:| ||:| |  
Db 1 MGDLSLTPGGSMGLQVNRGSQSSLEGAPATAPEP-HSLGILHASYSVSHRVRPWWDDITS 59

Qy 61 CQQKWDQRQILKDVSLYIESGQIMCILGSSSGSGKTTLLDAISGRLRRTGTLEGEVFNVC 120  
|:|:| |||||:|||||:||||| | || ||:| |  
Db 60 CRQQWTRQILKDVSLYVESGQIMCILGSSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRA 119

Qy 121 LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTELSLSH 180  
||:||||||| ||||| |||:|:| : : ||||| |||||  
Db 120 LRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSH 179

Qy 181 VADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240  
|||:|:|:|:| |||:|||||:|||||: |||||:| || |  
Db 180 VADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMMLFDEPTTGLDCMTANQIVLLVELA 239

Qy 241 RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCCGYPCPEHSNPF 300  
||:|||:|:||||||| |||||:|:|:| || | ||:|||||||  
Db 240 RRNRIVVLTIHQPRSELFQLFQDKIAILSFQELIFCGTPAEMLDFFNDCGYPCPEHSNPF 299

Qy 301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPVFKTK 360  
|||||:||||| ||||:| |:|:| | | |:||| :|:|||||||  
Db 300 FYMDLTSVDTQSKEREIETSKRVQMIESAYKSAICHKTLKNIERMKHLKTLPMPVFKTK 359

Qy 361 DPPGMFGKLGVLRLRVTRNLNRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNTLKGAVQDR 420  
| ||:| |||||:| || || |:|||||:|:|:|:| |||:| |  
Db 360 DSPGVFSKLGVLRLRVTRNLVRNKLAVITRLLQNLIMGLFLFFVLRVRSNVLKGAIQDR 419

Qy 421 VGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLPPFSVIAT 480  
||||| |||||:||||| ||||:| | |||||:| |  
Db 420 VGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLPPFSVVAT 479

Qy 481 VIFSSVCYWTGLYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI 540  
:|||||:|||||:|||||:|||||:|||||  
Db 480 MIFSSVCYWTGLHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI 539

Qy 541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPM 600  
:|:|:|:|:| |||: ||||| ||||| || | |: :| |  
Db 540 AGVLVGSGLRNIQEMPIPFKIISYFTFQKYCCEILVVNEFYGLNFTCGGSNVSVTTNPM 599

Qy 601 CAITQGVQFIEKTCPGATSREFTANFLILYGFIPALVILGIVIFKVRDYLSR 652  
 || |||:||||||| ||||| |||||:|:|:|  
 Db 600 CAFTQGIQFIEKTCPGATSREFTMNLILYSFIPALVILGIVVFKIRDHLISR 651

RESULT 10

AAU96992

ID AAU96992 standard; protein; 651 AA.

XX

AC AAU96992;

XX

DT 30-JUL-2002 (first entry)

XX

DE Human ABCG5 mutant E146Q protein sequence.

XX

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;  
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;  
 KW mutant; mutein.

XX

OS Homo sapiens.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Misc-difference 146

FT /note= "Wild-type Glu substituted by Gln"

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic  
 PT acid encoding the polypeptide, useful for treating sitosterolemia,  
 PT arteriosclerosis and heart diseases.

XX

PS Claim 12; Page; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette  
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a  
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart  
 CC disease. The molecules of the invention are also useful for identifying a  
 CC compound which alters ABCG5 activity level comprising contacting a cell  
 CC culture or mammal which have ABCG5 polypeptide with a compound and  
 CC measuring ABCG5 biological activity in the cell culture or in mammal,  
 CC where an increase or decrease in ABCG5 biological activity compared to  
 CC ABCG5 biological activity in a control cell culture or mammal not

CC contacted with the compound, identifies a compound that increases or  
CC decreases ABCG5 activity respectively. The cell culture or mammal  
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The  
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the  
CC polypeptide in a cell culture or mammal is also compared with that of a  
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.  
CC Stimulation of ABCG5 activity is useful for treating or preventing  
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's  
CC disease. The method of the invention is useful for increasing cholesterol  
CC excretion and/or decreasing cholesterol adsorption. The present amino  
CC acid sequence represents the human ABCG5 mutant E146Q protein of the  
CC invention. Note: This sequence is not shown in the specification but is  
CC derived from the wild-type human ABCG5 protein (AAU96984) given on pages  
CC 35-36 of the specification

XX

SQ Sequence 651 AA;

Query Match 81.4%; Score 2741.5; DB 5; Length 651;  
Best Local Similarity 80.1%; Pred. No. 1.7e-253;  
Matches 522; Conservative 65; Mismatches 64; Indels 1; Gaps 1;

Qy 1 MGELPFLSPEGARGPHINRGSLSSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS 60  
||:| |:| |:| :||| |||| | | ||||:| | ||||:| | ||:| |  
Db 1 MGDLSLTPGGSMGLQVNRGSQSSLEGAPATAPEP-HSLGILHASYSVSHRVRPWWIDITS 59

Qy 61 CQQKWDRQILKDVSLYIESGQIMCILGSSSGSGKTTLLDAISGRLRRTGTLEGEVFNVCCE 120  
|:|:| | ||||| |||||: ||||| |||||: ||||| | || | ||: |||  
Db 60 CRQQWTRQILKDVSLYVESGQIMCILGSSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRA 119

Qy 121 LRRDQFQDCFSYVLQSDVFLSSLTVRQTLRYTAMLALCRSSADFYNNKVEAVMTLSLSH 180  
|||: ||||| ||||| |||||: || |||: ||| | : : ||||| |||||  
Db 120 LRREQFQDCFSYVLQSDTLLSSLTVRQTLHYTALLAIRRGNPGSFQKKVEAVMAELSLSH 179

Qy 181 VADQMIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240  
|||: |||: |||: |||||: |||||: |||||: |||||: |||||: |||||  
Db 180 VADRLIGNYSLGGISTGERRRVSIQAQLLQDPKVMMLFDEPTTGLDCMTANQIVLLLELA 239

Qy 241 RRDRIVIVTIHQPRSELFQHFQDKIALITYGELVFCGTPPEMLGFFNCGYPCPEHSNPF 300  
||: |||: ||||| |||||: |||||: |||||: ||||| ||| |||: ||||| |||||  
Db 240 RRNRIVVLTIHQPRSELFQLFQDKIALISFGELIFCGTPAEMLDFFNDCGYPCPEHSNPF 299

Qy 301 FYMDLTSVDTQSREREIETIKRVQMLECAFKESDIYHKILENIERARYLKTLPMPVFKTK 360  
|| ||||| |||||: ||||| |||||: |||: ||| | || | ||||| : ||||| |||||  
Db 300 FYMDLTSVDTQSKEREIETSKRVQMIESAYKKAICHKTLKNIERMKHLKTLPMPVFKTK 359

Qy 361 DPPGMFGKLGVLRLRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNTLKGAVQDR 420  
| ||:| | ||||| |||||: ||| ||| ||: ||||| |||||: |||: |||: ||| |||||: |||  
Db 360 DSPGVFSKLGVLRLRVTRNLVRNKLAVITRLLQNLIMGLFLFFVLRVRSNVLKGAIQDR 419

Qy 421 VGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLFPFSVIAT 480  
|| |||| | ||||| |||||: ||||| |||||: ||||| |||||: ||||| |||||: |||  
Db 420 VGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLFPFSVAT 479

Qy 481 VIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPINVSIVALLSI 540  
: ||||| |||||: ||||| |||||: ||||| |||||: ||||| |||||: |||||  
Db 480 MIFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPINVSIVALLSI 539



CC culture or mammal which have ABCG5 polypeptide with a compound and  
 CC measuring ABCG5 biological activity in the cell culture or in mammal,  
 CC where an increase or decrease in ABCG5 biological activity compared to  
 CC ABCG5 biological activity in a control cell culture or mammal not  
 CC contacted with the compound, identifies a compound that increases or  
 CC decreases ABCG5 activity respectively. The cell culture or mammal  
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The  
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the  
 CC polypeptide in a cell culture or mammal is also compared with that of a  
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.  
 CC Stimulation of ABCG5 activity is useful for treating or preventing  
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's  
 CC disease. The method of the invention is useful for increasing cholesterol  
 CC excretion and/or decreasing cholesterol adsorption. The present amino  
 CC acid sequence represents the human ABCG5 mutant R389H protein of the  
 CC invention. Note: This sequence is not shown in the specification but is  
 CC derived from the wild-type human ABCG5 protein (AAU96984) given on pages  
 CC 35-36 of the specification

XX

SQ Sequence 651 AA;

Query Match 81.3%; Score 2739.5; DB 5; Length 651;  
 Best Local Similarity 80.1%; Pred. No. 2.6e-253;  
 Matches 522; Conservative 64; Mismatches 65; Indels 1; Gaps 1;

Qy	1	MGELPFLSPEGARGPHINRGSLSSEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS	60
		:   :   :  :                 :      :    :	
Db	1	MGDLSSLTPGGSMLQVNRGSQSSELEGAPATAPEP-HSLGILHASYSVSHRVRPWWDITS	59
Qy	61	CQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVC	120
		: :       : : : : : : : : : : : : : : : : : : :	
Db	60	CRQQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRA	119
Qy	121	LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTLSLSH	180
		:               : :    : :	
Db	120	LRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSH	179
Qy	181	VADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLLAELA	240
		: : : : : : : : : : : : : : : : : : :	
Db	180	VADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMMLFDEPTTGLDCMTANQIVLLVELA	239
Qy	241	RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPPEMLGFFNNGCYPCPEHSNPF	300
		:   : : : : : : : : : : : : : : : : : : :	
Db	240	RRNRIVVLTIHQPRSELFQLFQDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPF	299
Qy	301	FYMDLTSVDTQSREREIETYKRVQMLECAFESDIYHKILENIERARYLKTLPMPVFPKTK	360
		: : : : : : : : : : : : : : : : : : :	
Db	300	FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSAICHKTLKNIERMKHLKTLPMPVFPKTK	359
Qy	361	DPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNLTKGAVQDR	420
		:       : :        : : : : : : : : : : : :	
Db	360	DSPGVFSKLGVLRRVTRNLVRNKLAVITHLLQNLIMGLFLFFVLRVRSNVLKGAIQDR	419
Qy	421	VGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLFPFSVIAT	480
		: : : : : : : : : : : : : : : :	
Db	420	VGLLYQFVGATPYTGMLNAVNLFVRLAVSDQESQDGLYQKWQMLLAYALHVLFPFSVVAT	479

Qy	481	VIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFLLTVLLGIVQNPNIVNSIVALLSI	540
		:     :     :     :     :     :     :     :     :     :     :     :	
Db	480	MIFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLLTVLLGIVQNPNIVNSVVALLSI	539
Qy	541	SGLLIGSGFIRNIQEMPIPLKILGYFTTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHMP	600
		::  :    :     :     :     :     :     :     :     :     :     :	
Db	540	AGVLVGSGFLRNIQEMPIPFKIISYFTTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPM	599
Qy	601	CAITQGQVFIEKTCPGATSRETMNFILYGFIPALVILGIVIFKVRDYLIISR	652
		:    :     :     :     :     :     :     :     :     :	
Db	600	CAFTOGIQFIEKTCPGATSRETMNFLILYSFIPALVILGIVVFKIRDHLISR	651

AAU96989

ID AAU96989 standard; protein; 651 AA.

XX

AC AAU96989;

XX

DT 30-JUL-2002 (first entry)

XX

DE Human ABCG5 mutant R419H protein sequence.

XX

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;

KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;

KW mutant; mutein.

XX

OS Homo sapiens.

OS Synthetic.

XX

FH	Key	Location/Qualifiers
----	-----	---------------------

FT Misc-difference 419

```
FT      /note= "Wild-type Arg substituted by His"
```

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic

PT acid encoding the polypeptide, useful for treating sitosterolemia,

PT arteriosclerosis and heart diseases.

XX

PS Claim 9; Page; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette



CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a  
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart  
 CC disease. The molecules of the invention are also useful for identifying a  
 CC compound which alters ABCG5 activity level comprising contacting a cell  
 CC culture or mammal which have ABCG5 polypeptide with a compound and  
 CC measuring ABCG5 biological activity in the cell culture or in mammal,  
 CC where an increase or decrease in ABCG5 biological activity compared to  
 CC ABCG5 biological activity in a control cell culture or mammal not  
 CC contacted with the compound, identifies a compound that increases or  
 CC decreases ABCG5 activity respectively. The cell culture or mammal  
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The  
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the  
 CC polypeptide in a cell culture or mammal is also compared with that of a  
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.  
 CC Stimulation of ABCG5 activity is useful for treating or preventing  
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's  
 CC disease. The method of the invention is useful for increasing cholesterol  
 CC excretion and/or decreasing cholesterol adsorption. The present amino  
 CC acid sequence represents the human ABCG5 mutant R419H protein of the  
 CC invention. Note: This sequence is not shown in the specification but is  
 CC derived from the wild-type human ABCG5 protein (AAU96984) given on pages  
 CC 35-36 of the specification

XX

SQ Sequence 651 AA;

Query Match 81.3%; Score 2739.5; DB 5; Length 651;  
 Best Local Similarity 80.1%; Pred. No. 2.6e-253;  
 Matches 522; Conservative 64; Mismatches 65; Indels 1; Gaps 1;

Qy	1	MGELPFLSPEGARGPHINRGSLSSEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS	60
		:   :   :  :               :     :     :	
Db	1	MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEP-HSLGILHASYSVSHRVRPWWDITS	59
Qy	61	CQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVC	120
		: :     :     :     :     :             :	
Db	60	CRQQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRA	119
Qy	121	LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTLSLSH	180
		:                 : :   : :	
Db	120	LRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSH	179
Qy	181	VADQMIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA	240
		: : : :    :         :    :    :    :    :	
Db	180	VADRLIGNYSLGGISTGERRRVSIQAQLLQDPKVMMLFDEPTTGLDCMTANQIVVLLVELA	239
Qy	241	RRDRIVVTIHQPRSELFQHFQDKIAILTYGELVFCGTPPEMLGFFNNGCYPCPEHSNPF	300
		:   : :            : :            :	
Db	240	RRNRIVVLTIHQPRSELFQLFQDKIAILSFGEIFCGTPAEMLDFFNDCGYPCPEHSNPF	299
Qy	301	FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPF	360
		:         :   : :        :    : :	
Db	300	FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMPF	359
Qy	361	DPPGMFGKLGVLRRVTRNLNRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNTLKGAVQDR	420
		:             :            :    :    :    :	
Db	360	DSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDH	419

Qy	421 VGLLYQLVGATPYTGMLNAVNLFPM LRAVSDQESQDGLYHKWQM LLAYVLHVL PFSVIAT	480
	:           :           :	
Db	420 VGLLYQFVGATPYTGMLNAVNLF PVLRVSDQESQDGLYQKWQM LLAYALHVL PFSVVAT	479
Qy	481 VIFSSVCYWTLGLYPEVARFGYFSAALLAPHLIGEF LTVLLGIVQNPNIVNSIVALLSI	540
	:           :           :           :	
Db	480 MIFSSVCYWTLGLHPEVARFGYFSAALLAPHLIGEF LTVLLGIVQNPNIVNSSVVALLSI	539
Qy	541 SGLLIGSGFIRNIQE MPIPLKILGYFT FQKYCCEILV VNEFYGLNFTCGGSNTS MLNHMP	600
	:   :   :     :           :   :                           :   :	
Db	540 AGVLVSGSFLRNIQE MP I PFKIIS YFT FQKYCSEILV VNEFYGLNFTCGSSNV SVTTNPM	599
Qy	601 CAITQG VQFIEKTC PGATS RFTAN FLILYGFI PALVILGIV IFKVRDY LISR	652
	:                             :   :   :	
Db	600 CAFTQGI QFIEKTC PGATS RFTM N FLILYSFI PALVILGIV VF KIRDHLISR	651

XX  
PS Claim 10; Page; 66pp; English.

XX  
CC The present invention relates to a new mammalian ATP-binding cassette  
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a  
CC predisposition for developing sitosterolemia, arteriosclerosis or heart  
CC disease. The molecules of the invention are also useful for identifying a  
CC compound which alters ABCG5 activity level comprising contacting a cell  
CC culture or mammal which have ABCG5 polypeptide with a compound and  
CC measuring ABCG5 biological activity in the cell culture or in mammal,  
CC where an increase or decrease in ABCG5 biological activity compared to  
CC ABCG5 biological activity in a control cell culture or mammal not  
CC contacted with the compound, identifies a compound that increases or  
CC decreases ABCG5 activity respectively. The cell culture or mammal  
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The  
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the  
CC polypeptide in a cell culture or mammal is also compared with that of a  
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.  
CC Stimulation of ABCG5 activity is useful for treating or preventing  
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's  
CC disease. The method of the invention is useful for increasing cholesterol  
CC excretion and/or decreasing cholesterol adsorption. The present amino  
CC acid sequence represents the human ABCG5 mutant R419P protein of the  
CC invention. Note: This sequence is not shown in the specification but is  
CC derived from the wild-type human ABCG5 protein (AAU96984) given on pages  
CC 35-36 of the specification

XX  
SQ Sequence 651 AA;

Query Match 81.3%; Score 2737.5; DB 5; Length 651;  
Best Local Similarity 80.1%; Pred. No. 4e-253;  
Matches 522; Conservative 64; Mismatches 65; Indels 1; Gaps 1;

```
Qy      1 MGELPFLSPGARGPHINRGSLSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS 60
      ||:| |:| |:| :||| ||| | | |||:| | |||:| | ||:| |
Db      1 MGDLSLTPGGSMGLQVNRGSQSSLEGAPATAPEP-HSLGILHASYSVSHRVRPWWDITS 59

Qy     61 CQQKWDRQILKDVSLYIESGQIMCILGSSSGSKTTLDDAISGRLRRTGTLEGEVFNVC 120
      |:|:| |||||:|||||:|||||:||||| | || ||:| |
Db     60 CRQQWTRQILKDVSLYVESGQIMCILGSSSGSKTTLDDAMSGRLGRAGTFLGEVYVNGRA 119

Qy    121 LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKEAVMTLSLSH 180
      |||:||||| ||||| ||||| |||:|:| | : | ||||| |||||
Db    120 LRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRGNPGSFQKKVEAVMAELSLSH 179

Qy    181 VADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240
      |||:|:|:|:| ||||:|||||:|||||: ||||| |||||:| | ||
Db    180 VADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMMLFDEPTTGLDCMTANQIVVLLVELA 239

Qy    241 RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCGYPCPEHSNPF 300
      ||:|||:|:||||| |||||:|:|:| ||| |||:||||| |||||
Db    240 RRNRIVVLTIHQPRSELFQLFQDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPF 299

Qy    301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPVFKTK 360
      |||||:||||| ||||:| |:|:| | || |:||| |:||||| |||||
Db    300 FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMPVFKTK 359
```

Qy	361	DPPGMFGKLGVLRRVTRNLNMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNTTLKGAVQDR	420
Db	360	DSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDP	419
Qy	421	VGLLYQLVGATPYTGMLNAVNLFPM LRAVSDQESQDGLYHKWQMLLAYVLHVL PFSVIAT	480
Db	420	VGLLYQFVGATPYTGMLNAVNLF PVLRAVSDQESQDGLYQKWQMLLAYALHVL PFSVVAT	479
Qy	481	VIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEF LTLVLLGIVQNPNI VNSIVALLSI	540
Db	480	MIFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEF LTLVLLGIVQNPNI VNSV VALLSI	539
Qy	541	SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILV VNEFYGLNFTCGGSNTSMLNHPM	600
Db	540	AGVLVSGSGLRNIQEMPIPFKIISYFTFQKYCSEILV VNEFYGLNFTCGGSNVSVTTNPM	599
Qy	601	CAITQGVQFIEKTCPGATSRTANFLILYGFIPALVILGIVIFKVRDY LISR	652
Db	600	CAFTOGIQFIEKTCPGATSRTMNFILYLSFIPALVILGIVVFKIRDHLISR	651

CC of ATP-binding cassette (ABC) transporters by influencing the  
 CC dimerisation of the nucleotide binding domains comprises using: (a) a  
 CC polypeptide (polyP) consisting of 5-50 amino acids comprising the D loop  
 CC sequence of an ABC transporter (ABP52049 to ABP52091); (b) a polyP  
 CC consisting of the D loop sequence of an ABC transporter; (c) a peptide  
 CC mimetic or antisense peptide of (a) or (b). ABC transporters have  
 CC antibacterial, fungicide and protozoacide activities. (M1) is useful for  
 CC selectively modulating the activity of ABC transporters belonging to the  
 CC group of multidrug transporter/P-glycoproteins. Bacterial, fungal or  
 CC protozoal ABC transporters are involved in the infection of a mammal or  
 CC in the induction of resistance to antibiotics or drugs in a mammal. (M1)  
 CC is useful for preventing, treating or alleviating diseases associated  
 CC with functionality of an ABC transporter. ABP52092 to ABP52140 represent  
 CC ABC transporter proteins given in the exemplification of the present  
 CC invention

XX

SQ Sequence 649 AA;

Query Match 80.8%; Score 2722.5; DB 5; Length 649;  
 Best Local Similarity 79.9%; Pred. No. 1.1e-251;  
 Matches 521; Conservative 64; Mismatches 64; Indels 3; Gaps 2;

Qy	1	MGELPFLSPEGARGPHINRGSLSSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS	60
		:     :     :     :	
Db	1	MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEP-HSLGILHASYSVSHRVRPWWDITS	59
Qy	61	CQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVC	120
		: :	
Db	60	CRQQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRA	119
Qy	121	LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTLSLSH	180
		:	
Db	120	LRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRGNPGSFQKKVEAVMAELSLSH	179
Qy	181	VADQMIGSYNFGGISSGERRRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA	240
		: :     :       :	
Db	180	VADRLIGNYSLGGISTGERRRRVSIAAQLLQDPKVMMLF--PTTGLDCMTANQIVVLLVEIA	237
Qy	241	RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPF	300
		:       :	
Db	238	RRNRIVVLTIHQPRSELFQLFQDKIAILSFGEIIFCGTPAEMLDFFNDCGYPCPEHSNPF	297
Qy	301	FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPVPEKTK	360
		:	
Db	298	FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMPVPEKTK	357
Qy	361	DPPGMFGKLGVLRRVTRNLNRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNTLKGAVQDR	420
		:                     :                     :   :     :	
Db	358	DSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDR	417
Qy	421	VGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLFPFSVIAT	480
		:	
Db	418	VGLLYQFVGATPYTGMLNAVNLFVLRVAVSDQESQDGLYQKWQMLLAYALHVLFPFSVVAT	477
Qy	481	VIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPNIIVNSIVALLSI	540
		:                     :	
Db	478	MIFSSVCYWTGLHPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPNIIVNSVALLSI	537

[illegible]

AAU96991

XX

XX

XX

XX

XX

OS Synthetic.

FT

XX

XX

XX

XX

XX

PA

PA

XX

XX

XX

XX

XX

CC compound which alters ABCG5 activity level comprising contacting a cell  
 CC culture or mammal which have ABCG5 polypeptide with a compound and  
 CC measuring ABCG5 biological activity in the cell culture or in mammal,  
 CC where an increase or decrease in ABCG5 biological activity compared to  
 CC ABCG5 biological activity in a control cell culture or mammal not  
 CC contacted with the compound, identifies a compound that increases or  
 CC decreases ABCG5 activity respectively. The cell culture or mammal  
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The  
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the  
 CC polypeptide in a cell culture or mammal is also compared with that of a  
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.  
 CC Stimulation of ABCG5 activity is useful for treating or preventing  
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's  
 CC disease. The method of the invention is useful for increasing cholesterol  
 CC excretion and/or decreasing cholesterol adsorption. The present amino  
 CC acid sequence represents the human ABCG5 mutant R408X protein of the  
 CC invention. Note: This sequence is not shown in the specification but is  
 CC derived from the wild-type human ABCG5 protein (AAU96984) given on pages  
 CC 35-36 of the specification  
 XX  
 SQ Sequence 408 AA;

Qy	1	MGELPFLSPEGARGPHINRGSLSLSEQGSVTGTEARHSLGVLHVSYSVSNRVPWWNIKS	60
Db	1	MGDLSSLTTPGGSMGLQVNRGSQSSLEGAPATAPEP-HSLGILHASYSVSHRVRPWWDITS	59
Qy	61	CQQKWDRQILKDVSLYIESGQIMCILGSSSGSGKTTLLDAISGRLLRRTGTLEGEVFNVC	120
Db	60	CRQQWTRQILKDVSLYVESGQIMCILGSSSGSGKTTLLDAMSGRLGRAGTFLGGEVYVNGRA	119
Qy	121	LRRDQFQDCFSYVLQSDVFLSSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTLSLSH	180
Db	120	LRREQFQDCFSYVLQSDTLLSSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSH	179
Qy	181	VADQMIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLLAELA	240
Db	180	VADRIGNYSLGGISTGERRRVSIQAQLLQDPKVMMLFDEPTTGLDCMTANQIVVLLVELA	239
Qy	241	RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPPEMLGFFNNCGYPCPEHSNPF	300
Db	240	RRNRIVVLTIHQPRSELFQLFQDKIAILSFGEILFCGTPAEMLDFFNDCGYPCPEHSNPF	299
Qy	301	FYMDLTSVDTQSREREIETYSKRVQMLECAFKESDIYHKILENIERARYLKTLPMPVFFKTK	360
Db	300	FYMDLTSVDTQSKEREIETYSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFFKTK	359
Qy	361	DPPGMFGKLGVLRLRRVTRNLNRNKQAVIMRLVQNLIMGLFLIFYLLRVQ	409
Db	360	DSPGVFSKLGVLRLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVR	408

OM protein - protein search, using sw model

Run on: February 27, 2004, 07:11:48 ; Search time 14.7734 Seconds  
(without alignments)  
2278.426 Million cell updates/sec

Title: US-09-989-981A-2  
Perfect score: 3369  
Sequence: 1 MGELPFPLSPEGARGPHINRG.....PALVILGIVIFKVRDYLI SR 652

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Query		Length	DB	ID	Description
		Match					
1	693.5	20.6		655	4	US-09-245-808-1	Sequence 1, Appli
2	689.5	20.5		655	4	US-09-767-594-1	Sequence 1, Appli
3	457.5	13.6		1296	4	US-09-614-912-140	Sequence 140, App
4	419.5	12.5		617	4	US-09-614-912-138	Sequence 138, App
5	330.5	9.8		539	4	US-09-614-912-144	Sequence 144, App
6	262.5	7.8		384	4	US-09-489-039A-9127	Sequence 9127, Ap
7	259.5	7.7		1684	3	US-08-665-259-25	Sequence 25, Appl
8	259.5	7.7		1684	3	US-08-762-500-25	Sequence 25, Appl
9	259.5	7.7		1704	3	US-08-762-500-75	Sequence 75, Appl
10	247	7.3		653	4	US-09-543-681A-5411	Sequence 5411, Ap
11	246.5	7.3		210	4	US-09-543-681A-8215	Sequence 8215, Ap



12	246	7.3	373	4	US-09-543-681A-7638	Sequence 7638, Ap
13	243	7.2	242	4	US-09-134-001C-3832	Sequence 3832, Ap
14	242	7.2	1280	4	US-09-672-810-2	Sequence 2, Appli
15	242	7.2	1283	4	US-09-672-810-4	Sequence 4, Appli
16	240	7.1	229	4	US-09-134-000C-3584	Sequence 3584, Ap
17	240	7.1	406	4	US-09-489-039A-10003	Sequence 10003, A
18	238.5	7.1	248	4	US-09-134-001C-3731	Sequence 3731, Ap
19	238.5	7.1	344	4	US-09-489-039A-13987	Sequence 13987, A
20	237	7.0	402	4	US-09-107-532A-5360	Sequence 5360, Ap
21	236.5	7.0	1280	2	US-08-752-447-2	Sequence 2, Appli
22	236.5	7.0	1280	4	US-09-316-167-2	Sequence 2, Appli
23	236.5	7.0	1280	4	US-09-397-233-2	Sequence 2, Appli
24	236	7.0	329	4	US-09-107-532A-4844	Sequence 4844, Ap
25	236	7.0	1279	2	US-08-784-649A-2	Sequence 2, Appli
26	236	7.0	1279	4	US-09-672-810-6	Sequence 6, Appli
27	236	7.0	1280	2	US-08-583-276-19	Sequence 19, Appl
28	236	7.0	1280	4	US-09-767-594-2	Sequence 2, Appli
29	236	7.0	1280	4	US-09-672-810-5	Sequence 5, Appli
30	236	7.0	1280	6	5206352-4	Patent No. 5206352
31	235.5	7.0	358	4	US-09-489-039A-7399	Sequence 7399, Ap
32	235	7.0	573	4	US-09-489-039A-12091	Sequence 12091, A
33	234.5	7.0	1479	2	US-08-951-912-4	Sequence 4, Appli
34	234.5	7.0	1479	4	US-09-174-077-4	Sequence 4, Appli
35	234	6.9	1476	3	US-09-256-703-2	Sequence 2, Appli
36	234	6.9	1480	1	US-07-637-621-2	Sequence 2, Appli
37	234	6.9	1480	2	US-08-951-912-2	Sequence 2, Appli
38	234	6.9	1480	3	US-08-681-838A-2	Sequence 2, Appli
39	234	6.9	1480	3	US-08-681-838A-3	Sequence 3, Appli
40	234	6.9	1480	4	US-09-174-077-2	Sequence 2, Appli
41	234	6.9	1480	6	5240846-5	Patent No. 5240846
42	233.5	6.9	250	4	US-09-328-352-7153	Sequence 7153, Ap
43	233	6.9	476	4	US-09-107-532A-5247	Sequence 5247, Ap
44	233	6.9	1480	4	US-09-425-453A-2	Sequence 2, Appli
45	233	6.9	1480	4	US-09-425-453A-12	Sequence 12, Appl

#### ALIGNMENTS

```

RESULT 1
US-09-245-808-1
; Sequence 1, Application US/09245808
; Patent No. 6313277
; GENERAL INFORMATION:
; APPLICANT: Doyle, L. Austin
; APPLICANT: Abruzzo, Lynne V.
; APPLICANT: Ross, Douglas D.
; TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which
; TITLE OF INVENTION: encodes it
; FILE REFERENCE: Ross UMb conversion
; CURRENT APPLICATION NUMBER: US/09/245,808
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/073763
; EARLIER FILING DATE: 1998-02-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1

```

; LENGTH: 655  
; TYPE: PRT  
; ORGANISM: Human MCF-7/AdrVp cells  
US-09-245-808-1

Query Match 20.6%; Score 693.5; DB 4; Length 655;  
Best Local Similarity 29.0%; Pred. No. 3.9e-61;  
Matches 181; Conservative 142; Mismatches 246; Indels 55; Gaps 16;

```
Qy      25 LEQGSVTGTEARHS-----LGVLHVSYSVSNRVGPWWNIKSCQQKWDRQILKDV 73
      : ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      12 VSQGNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLKSG----FLPCRKPVEKEILSNI 67

Qy      74 SLYIESGQIMCILGSSSGSKTTLDDAISGRLLRRTGTLEGEVVFVNGCELRRDQFQDCFSYV 133
      : :: | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      68 NGIMKPG-LNAILGPTGGGKSSLLDVLAAARKDPSG-LSGDVLINGAP-RPANFKCNSGYV 124

Qy     134 LQSDVFLSSLTVRETLRYTAMLALCRSSADF-YNKKVEAVMTELSLSHVADQMIGSYNFG 192
      : | | | : : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     125 VQDDVVMGTLTVRENLQFSAAALRLATMTNHEKNERINRVIQELGLDKVADSKVGTQFIR 184

Qy     193 GISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQ 252
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     185 GVSggerKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKRMSKQGRTIIFSIIHQ 244

Qy     253 PRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCGYPCPEHSNPFDFYMDLTSVDTQ- 311
      | | : | : | | : | | | | | | | | | | | | | | | | | | | | | | |
Db     245 PRYSIFKLFDLTLASGRLMFHGPAQEALGYFESAGYHCEAYNNPADFFLDIINGDSTA 304

Qy     312 -SREREIETYKRVQMLECAFKESDIYHKI-----LENIERARYLKT 351
      : | | | : | : : | : : : | : | | | | | | | | | | | | | | | | |
Db     305 VALNRE-EDFKATEIIIEPSKQDKPLIEKLAIEIYNSSFYKETKAELHQLSGGEKKKKITV 363

Qy     352 LPMVPFKTKDPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLMGLFLIFYLLRVQNN 411
      : : | | : | : | : | | | | | | | | | | | | | | | | | | | | | |
Db     364 FKEISYTT----SFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 419

Qy     412 TLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMRLRAVSDQESQDGLYHKWQMLLAYVL- 470
      : : : | | | | | : : : | | | | | | | | | | | | | | | | | |
Db     420 ST--GIQNRAGVLFFLTNQCFS--VSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLS 476

Qy     471 HVLPFVSVIATVIFSSVCYWTGLYPEVARFGYFSAALLAPHLIGEFLLTVLLGIVQNPNI 530
      : | | : : : | | : | | | | | | | | | | | | | | | | | | | | | |
Db     477 DLLPMTMLPSIIFTCIVYFMLGLKPKADAFFVMMFTLM---MVAYSASSMALAIAAGQSV 533

Qy     531 VNSIVALLSIS--GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTC 588
      | : | | : | : | | : | : | | | | | | | | | | | | | | | | | | |
Db     534 VSVATLLMTICFVFMIFSGLLVNLTIASWLSWLQYFSIPRYGFTALQHNEFLGQNF-C 592

Qy     589 GGSNTSMLNHPMCAITQGVQFIEK 612
      | | : | | | | : : |
Db     593 PGLNATGNNPCNYATCTGEEYLVK 616
```

RESULT 2  
US-09-767-594-1  
; Sequence 1, Application US/09767594



```

      :   :|:| |:|: |   : : :| | | : : : | | |   | :|
Db      420 ST--GIQNRAGVLFFLT TNQCFSS--VSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLS 476

Qy      471 HVL PFSVIATVIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEF LTLVLLGIVQNPNI 530
      :| | : : :| | : : | : | | | | : | : : : : | | : :
Db      477 DLLPMRMLPSIIFTCIVYFMLGLKPKADAFFVMMFTLM--MVAYSASSMALAIAAGQSV 533

Qy      531 VNSIVALLSIS--GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTC 588
      | :   | : : |   : : | | : | : : | | | | : : | | | | |
Db      534 VSVATLLMTICFVFM MIFSGLLVNLTTIASWLSWLQYFSIPRYGFTALQHNEFLGQNF-C 592

Qy      589 GGSNTSMLNHPMCAITQGVQFIEK 612
      | | : | | | : : : |
Db      593 PGLNATGNNPCNYATCTGEEYLVK 616

```

# RESULT 3

US-09-614-912-140

```

; Sequence 140, Application US/09614912
; Patent No. 6677502
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Rafalski, Antoni
; APPLICANT: Orozco, Buddy
; APPLICANT: Miao, Gou-Hau
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Lee, Jian Ming
; APPLICANT: Sakai, Hajime
; APPLICANT: Weng, Zude
; APPLICANT: Caimi, Perry G
; APPLICANT: Anderson, Shawn
; TITLE OF INVENTION: Plant Metabolism Genes
; FILE REFERENCE: BB1378 US NA
; CURRENT APPLICATION NUMBER: US/09/614,912
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: 60/143,401
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/143,412
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/146,650
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/170,906
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/172,959
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/172,946
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 140
; LENGTH: 1296
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-614-912-140

```

```

Query Match          13.6%;  Score 457.5;  DB 4;  Length 1296;
Best Local Similarity 26.8%;  Pred. No. 1.1e-36;

```

Matches 169; Conservative 100; Mismatches 244; Indels 117; Gaps 23;

```
Qy      85 ILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVCCELRRDQFQDCFSYVLQSDVFLSSLT 144
      :||  ||| ||| |::|| :   ||| || ||   |  :: ||| :  :|
Db      9  LLGPPSSGKTTLLLAGLKDPSLRGGEVTVNGFELEEFVAQKTAAYISQTDVHVGMT 68

Qy     145 VRETLRYTAMLALCRSSADFYNNKKEAVMTE----- 175
      |:||| ::|   | :   | :   |
Db      69 VKETLDFSAR---CQGVGTYKDLLTELARREKEAGIRPEPEVDLFMKATSMEGVESSLQT 125

Qy     176 -----LSLSHVADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTA 229
      | |   || ::|   ||| |::||:   ::   ||: ||| :||| |
Db     126 DYTLRILGLDICADTIVGDMQQRGISGGQKKRVTTGEMIVGPTKVLFMDEISTGLDSSTT 185

Qy     230 NQIVLLLLAELAR-RDRIVIVTIHQPRSELFQHFDKIAILTYGELVFCGTPEEMLGFFNNC 288
      ||| | ::   :  :::: || | :| || | :| :| :| | :| | :|
Db     186 FQIVKCLQQIVHLGEATILMSLLQPAETFEFDDIILLSEGQIVYQGPREYVLEFFESC 245

Qy     289 GYPCPEHSNPFDFYMDLTSVDTQSR--EREIETYKRVQMLECA--FKESDIYHKILENIE 344
      |: |||   || ::|| | :   :  | : : : | | ||   :| |:
Db     246 GFRCPERKGTADFLQEVTSKKDQEQYWADKHRPYRYSVSEFAQRFRK---FHVGLQ--- 299

Qy     345 RARYLKTLPMVPF-KTKDPPG--MFGKLGVLRRVTRN-----LMRNKQAVIMRLVQ 393
      | :   ||| ||:   :| | |   : :   : ||   | : :|
Db     300 ----LENHLSVPFDKTRSHQAALVFSKQSVSTTELLKASFKEWLLIKRNSFVYIFKTIQ 355

Qy     394 NLIMGLFLIFYLLRVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQ- 452
      ::| |   || | :|   : | | :| || :: ::| | |   :
Db     356 LIIVALVASTVFLRTQMHTRN--LDD--GFVY--IGALLFSLIVNMFNGFAELSLTITRL 409

Qy     453 ----ESQDGL-YHKWQMLLAYVLHVLFPFSVIATVIFSSVCYWTGLGYPEVARFGYFSAAL 507
      : :| | | |   | | :  :|||:| :::: | |:|:| || || | |
Db     410 PVFFKHRDLLFYPAWIFTLPNVILRIPFSIIESIVWVIVTYTYTIGFAPEADRF--FKQLL 467

Qy     508 LAPHLIGEFLLTLVLLG-----IVQNPNIIVNSIVALLSISGLLIGSGFIRNIQ 554
      |   ||   : |   | :   ::   : | | :   ||
Db     468 LV-----FLIQQMAGGLFRATAGLCRSMIIAQTGALALLIFFVLGGFLLPKAFIPK-- 519

Qy     555 EMPIPLKILGYFTFQ-KYCCEILVVNEFYGLNFTCGGSNTSMLNHPMCAITQGVQFIEKT 613
      | ||: |   | |||| :   | :|::   | : :|
Db     520 ----WWIWGYWVSPLMYGYNALAVNEFYSPRW----MNKFVLDNNGVPKRLGIALME-- 568

Qy     614 CPGATSRTANFLILYGFIPALVILGIVIF 643
      ||   | :   :| | :|| :|
Db     569 --GANIFTDKNWF----WIGAAGLLGFTMF 592
```

#### RESULT 4

US-09-614-912-138

; Sequence 138, Application US/09614912

; Patent No. 6677502

; GENERAL INFORMATION:

; APPLICANT: Allen, Steve

; APPLICANT: Rafalski, Antoni

; APPLICANT: Orozco, Buddy

; APPLICANT: Miao, Gou-Hau

; APPLICANT: Famodu, Omolayo O.

```

; APPLICANT: Lee, Jian Ming
; APPLICANT: Sakai, Hajime
; APPLICANT: Weng, Zude
; APPLICANT: Caimi, Perry G
; APPLICANT: Anderson, Shawn
; TITLE OF INVENTION: Plant Metabolism Genes
; FILE REFERENCE: BB1378 US NA
; CURRENT APPLICATION NUMBER: US/09/614,912
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: 60/143,401
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/143,412
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/146,650
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/170,906
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/172,959
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/172,946
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 138
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Zea mays
US-09-614-912-138

```

```

Query Match          12.5%; Score 419.5; DB 4; Length 617;
Best Local Similarity 24.6%; Pred. No. 2.3e-33;
Matches 152; Conservative 130; Mismatches 240; Indels 97; Gaps 25;

```

```

Qy      66 DR-QILKDVSLYIESGQIMCILGSSSGSKTTLLDAISGRLLRRTGTLEGEVFNVCCELRRD 124
      || |:|::|: | : :|| ||:|||||:| :|| : | :||:: : | :
Db      34 DRLQLLREVTGSFRPGVLTALMGVSGAGKTTLMVDVLAGR-KTGGYIEGDIRIAGYPKNQA 92

Qy      125 QFQDCFSYVLQSDVFLSSLTVRETLRYTAMLAL-----CRSSADFYNKKVEAVMTELSL 178
      |      | |:|: :|||:| |:| | | : | : |:| | : |
Db      93 TFARISGYCEQNDIHSPPQTVRESLIYSAFLRLPGKIGDQEITDDIKMQFVDEVMEVEL 152

Qy      179 SHVADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLA 238
      :: | ::| |:| :|:|:| | :| :| :: :|||:| | | : :
Db      153 DNLRLDALVGLPGITGLSTEQRKRLTIAVELVANPSIIFMDEPTSGLDARAAAIVMRTVRN 212

Qy      239 LARRDRIVVTIHQPRSELFQHFQDKIAILTY-GELVFCG----TPEEMLGFFNNC-GYP- 291
      | |: | ||| :|:| ||:: :| |::: | :|: :| | |
Db      213 TVDTGRTVVCTIHQPSIDIFESFDELLLLKRGQVIYSGKLGRNSQKMVEYFEAIPGVPK 272

Qy      292 CPEHSNPFDFYMDLTSVDTQSR-EREIETYKRVQMLECAFKESDIY--HKILENIERARY 348
      : || : :::|| |: | : : | :| ||:| :|:| | : :
Db      273 IKDKYNPATWMLEVSSVATEVRLKMDFAKY-----YETSDLYKQNKVLVN-QLSQP 322

Qy      349 LKTLPMVPFKTKDPPGMFGKLGVLRLRRVTRNLMRNKQAVIMRLVQNLIMGLFL--IFYLL 406
      : | |: | : | : | : :| |:| | | ||: :
Db      323 EPGTSDLYFPTEYSQSTIGQFKACLWKQWLTYWRSPDYNLVRYSFTELLVALLLGSIFWRI 382

```

Qy	407	--RVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPML---	RAVSDQESQDGLYHK	461
		: ::       :   :   :   :   :   :   :		
Db	383	GTNMEDATTLGMV---IGAMYT---AVMFIGINNCS	TVQPVVSIERTV	FYRERAAAGMYSA 436
Qy	462	WQMLLAYVLHVLPPFSVIATVIFSSVCY-----	WTLGLYPEVARFGYFSAALLAPHLIGE	515
		:     : :   : :     :		
Db	437	MPYAIAQVVIPIPVFVQTTYTTLIVYAMMSFQWT	AVKFFWFFFI	SYFS----- 485
Qy	516	FLTLVLLGIVQ---NPN-IVNSIVALLSISGLLIGSG	FIRNIQEMPIPLKILGYFTFQKY	571
		: :         :     :     : : : :		
Db	486	FLYFTYYGMMAVSISPNEHVASIFAAAFSLFNLFSG	FF-----IPRP-RIPGWWIWYYW	539
Qy	572	CCEILVVNEFYGLNFTCGGSNTSMLNHPMCAITQGV	QFIEKTCPGATS	SRFTANFLILYGF 631
		:         : :     : : :		
Db	540	ICPLAWT--VYGLIVTQYGDLEDLISVP-----	GESEQTISYYVTHHF	580
Qy	632	-----IPALVILGI		640
		: :		
Db	581	GYHRDFLPVIAPVLVLEFAV		599

US-09-614-912-144

; Patent No. 6677502

; GENERAL INFORMATION:

; APPLICANT: Allen, Steve

; APPLICANT: Rafalski, Antoni

; APPLICANT: Orozco, Buddy

; APPLICANT: Miao, Gou-Hau

; APPLICANT: Famodu, Omolayo O.

; APPLICANT: Lee, Jian Ming

; APPLICANT: Sakai, Hajime

; APPLICANT: Weng, Zude

; APPLICANT: Caimi, Perry G

; APPLICANT: Anderson, Shawn

; TITLE OF INVENTION: Plant Metabolism Genes

; FILE REFERENCE: BB1378 US NA

; CURRENT APPLICATION NUMBER: US/09/614,912

; CURRENT FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: 60/143,401

; PRIOR FILING DATE: 1999-07-12

; PRIOR APPLICATION NUMBER: 60/143,412

; PRIOR FILING DATE: 1999-07-12

; PRIOR APPLICATION NUMBER: 60/146,650

; PRIOR FILING DATE: 1999-07-30

; PRIOR APPLICATION NUMBER: 60/170,906

; PRIOR FILING DATE: 1999-12-15

; PRIOR APPLICATION NUMBER: 60/172,959

: PRIOR FILING DATE: 1999-12-21

: PRIOR APPLICATION NUMBER: 60/172,946

: PRIOR FILING DATE: 1999-12-21

; NUMBER OF SEO ID NOS: 204

```

: SOFTWARE: Microsoft Office 97

```

: SEO ID NO 144

; LENGTH: 539

; TYPE: PRT

; ORGANISM: Triticum aestivum  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (272)..(273)  
US-09-614-912-144

Query Match 9.8%; Score 330.5; DB 4; Length 539;  
Best Local Similarity 21.9%; Pred. No. 2e-24;  
Matches 111; Conservative 120; Mismatches 210; Indels 67; Gaps 17;

```
Qy      108 GTLEGEVFNVCCELRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNK 167
      | :|||: |:|   ::: |   | |:|:   |:| |:| ::| | | : | :
Db      2  GYIEGEITVSGYPKKQETFARISGYCEQNDIHSPhVTIYESLVFSAWLRL-PAEVDSE 60

Qy      168 K--VEAVMTELSLSHVADQMIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLDEPTTGLD 225
      | :| :| : |: :   ::|   |::: |:|::| | |:| :| :: :|||:| | |
Db      61 KMFIEEIMDLVELTSLRGALVGLPGVNGLSTEQRKRLTIAVELVANPSIIFMDEPTSGLD 120

Qy      226 CMTANQIVLLLAELARRDRIVIVTIHQPRSELFQHFQDKIAILTY-GELVFCG----TPEE 280
      | :: :   | | : |||| | ::| : |::: ::   || :: |
Db      121 ARAAAIVMRTVRNTVNTGRTVCTIHQPSIDIFEAFDELFLMKRGGEIYVGPVGQNSAN 180

Qy      281 MLGFFNNC-----GYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRVQMLECAFKE 332
      :: :|   ||   || : ::::|   | : |:   : | ::
Db      181 LIEYFEEIEGISKIKDGY-----NPATWMLEVSS----SAQEEM---LGIDFAE-VYRQ 226

Qy      333 SDIYHKILENIERARYLKTLP-----VPFKTKDPPGMFGKLGVLRLRRVTRNLMRNKQ 385
      |::| : |   :| | |   : | | :   : | : : | |
Db      227 SELYQRNKE-----LIKELSMAPAGSSDLNFPTQYSRSEFTQCLACLWKQXXSYWRNPS 280

Qy      386 AVIMRLVQNLIMGLFLIFYLLRVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFP 445
      :||: ::|   : : | :   ||   : : | | |:| :| :| :|
Db      281 YTAVRLFLTIVIALMFGTMFWDLGSKTRRS--QDLFNAMGSMYAAVLYIGVQNSGVSQPV 338

Qy      446 L---RAVSDQESQDGLYHKWQMLLAYVLHVLFPFSVIATVIFSSVCYWTGLGYPEVARFGY 502
      : | | :| |:| :   |   | : :: :|: : | :|   ||:| :
Db      339 VVERTVIFYRERAAAGMYSAFPYAFGQVAIEFPYVLVQALIYGGVLVYSMIGFEWTVAKFLW 398

Qy      503 FSAALLAPHLIGEFLLTLVLLGIVQNPNIIVNSIVALLSISGLLIG---SGFIRNIQEMPIP 559
      : : | | :::|: |   || |:| : :   ||:: :| :|
Db      399 YLFFMYFTMLYFTFYGMMAVGLTPN----ESIAAISSAFYNVWNLFSGYLIPRPKLPI- 453

Qy      560 LKILGYFTFQKYCCEI-----LVVNEF 581
      :: : : | :   || ::|
Db      454 -----WWRWYSWICPVAWTLYGLVASQF 476
```

RESULT 6

US-09-489-039A-9127

; Sequence 9127, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
KLEBSIELLA

; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2004001



; CURRENT APPLICATION NUMBER: US/09/489,039A  
 ; CURRENT FILING DATE: 2000-01-27  
 ; PRIOR APPLICATION NUMBER: US 60/117,747  
 ; PRIOR FILING DATE: 1999-01-29  
 ; NUMBER OF SEQ ID NOS: 14342  
 ; SEQ ID NO 9127  
 ; LENGTH: 384  
 ; TYPE: PRT  
 ; ORGANISM: Klebsiella pneumoniae  
 US-09-489-039A-9127

Query Match 7.8%; Score 262.5; DB 4; Length 384;  
 Best Local Similarity 26.3%; Pred. No. 8.9e-18;  
 Matches 78; Conservative 58; Mismatches 118; Indels 43; Gaps 8;

Qy 68 QILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLLRRTGTLEGEVFNVCGLRRDQFQ 127  
 |:| |:| | |::: |:| | | | | | | |:| :| | : |:| :  
 Db 36 QVLNDISLDIPSGQMVALLGPSGSGKTTLLRIIAGLEHQT---SGHIRFHGTDVSRMHAR 92  
 Qy 128 D-CFSYVLQSDVFLSSLTVRETLRY--TAMLALCRSSADFYNNKVEAVMTELSLHVADQ 184  
 | :| | :| : : : | : | :| | : : : |:| :| :  
 Db 93 DRKVGFFVQHYALFRHMTVFDNIAFGLTVLPRRERPNAAIKAKVTKLLEMVQLAHLADR 152  
 Qy 185 MIGSYNFGGISSGERRRVSIQAQLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDR 244  
 :| |::|::| | :|:::| | | | : : | :| :  
 Db 153 YPAQ-----LSGGQKQQRVALARALAVEPQILLLLDEPFGALDAQVRKELRRWLRLHEELK 207  
 Qy 245 IVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPE-----EMLGFFNN----- 287  
 | : : | : |::: | : | | | :| |  
 Db 208 FTSVFVTHDQEEAMEVADRVMVMSQGNIEQADAPERVWREPSTRFVLEFMGEVNRLQGV 267  
 Qy 288 -----CGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRVQMLECAFK 331  
 || | : | | : : || || : : ||:| : |  
 Db 268 RGGQFHVGAHRWPLGY-TPAYQGFVDLFLRPWEVDI-SRRTSLDSPLPVQVLEASPK 322

# RESULT 7

US-08-665-259-25

; Sequence 25, Application US/08665259  
 ; Patent No. 6028173

## ; GENERAL INFORMATION:

; APPLICANT: Landes, Gregory M.  
 ; APPLICANT: Burn, Timothy C.  
 ; APPLICANT: Connors, Timothy D.  
 ; APPLICANT: Dackowski, William R.  
 ; APPLICANT: Van Raay, Terence J.  
 ; APPLICANT: Klinger, Katherine W.  
 ; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,  
 ; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME  
 ; NUMBER OF SEQUENCES: 73  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: GENZYME CORPORATION  
 ; STREET: One Mountain Road  
 ; CITY: Framingham  
 ; STATE: Massachusetts  
 ; COUNTRY: United States of America  
 ; ZIP: 01701

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/665,259
; FILING DATE: 17-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IG5-9.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1684 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-665-259-25

```

```

Query Match          7.7%; Score 259.5; DB 3; Length 1684;
Best Local Similarity 30.1%; Pred. No. 2.1e-16;
Matches 85; Conservative 48; Mismatches 102; Indels 47; Gaps 13;

```

```

Qy      30 VTGTEARHSLGLVHVSYSVSNRVGPWWNIKSCQKQWDRQILKDVSLEYIESGQIMCILGSS 89
      | | : : | | : | | : | | : | | : | | : | | : | | : | | : | | :
Db      507 VAGIKIKH----LSKVERVGNK-----DRAAVRDLNLNLNLYEGQITVLLGHN 548

Qy      90 GSGKTTLLDAISGRLRRTGTLEGEVVFVNGCELRRD--QFQDCFSYVLQSDVFLSSLTVRE 147
      | : | | | | : : | | : | | : | | : | | : | | : | | : | | :
Db      549 GAGKTTTTLMLTGLFPPT---SGRAYISGYEISQDMVQIRKSLGLCPQHDILFDNLTVAE 605

Qy      148 TLRYTAML-ALCRSSADFYNNKKVEAVMTELSLSHVADQMIGSYNFGGISSGERRRVSIAA 206
      | : | | | | : | | : | | : | | : | | : | | : | | : | | :
Db      606 HLYFYAQLKGLSR-----QKCPEEVKQMLHIIGLEDKWNSRSRF--LSGGMRRLKLSIGI 657

Qy      207 QLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIH-QPRSELFQHFDKIA 265
      | : | | : | | | | : | | : | | : | | : | | : | | : | | :
Db      658 ALIAGSKVLILDEPTSGMDAISRRAIWDL-QRQKSDRTIVLTTHFMDEADLLG--DRIA 714

Qy      266 ILTYGELVFCGTP---EEMLGFFNNCGYPC----PEHSNPF 300
      | : | | | | : : | | : | | : | | : | | : | | : | | :
Db      715 IMAKGELQCCGSSLFLKQKYG---AGYHMTLVKEPHCNPED 752

```

# RESULT 8

US-08-762-500-25

; Sequence 25, Application US/08762500

; Patent No. 6030806

; GENERAL INFORMATION:

; APPLICANT: Landes, Gregory M.

; APPLICANT: Burn, Timothy C.

; APPLICANT: Connors, Timothy D.

```

; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klinger, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/762,500
; FILING DATE: 09-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/665,259
; FILING DATE: 17-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10469
; FILING DATE: 17-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IG5-9.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1684 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-762-500-25

```

```

Query Match          7.7%; Score 259.5; DB 3; Length 1684;
Best Local Similarity 30.1%; Pred. No. 2.1e-16;
Matches 85; Conservative 48; Mismatches 102; Indels 47; Gaps 13;

```

```

Qy      30 VTGTEARHSLGVLHVSYSVSNRVGPWWNIKSCQKWDRQILKDVSLYIESGQIMCILGSS 89
      | | : : |   |   : | | :           || :|:|:| : | | | :| | :
Db      507 VAGIKIKH----LSKVFRVGNK-----DRAAVRDLNLNLYEGQITVLLGHN 548

Qy      90 GSGKTTLLDAISGRLRRTGTLEGEVFNVCCELRRD--QFQDCFSYVLQSDVFLSSLTVRE 147
      |:| || | | : : |   |   | : : | | : : | | :   | | :   :| | |
Db      549 GAGKTTTTLTMLTGLFPPT---SGRAYISGYEISQDMVQIRKSLGLCPQHDILFDNLTVAE 605

Qy      148 TLRYTAML-ALCRSSADFYNNKKVEAVMTLSLSHVADQMIGSYNFGGIISSGERRRVSIAA 206
      | : | | | |   |   | | |   | : : | :   | : | | | | :| |

```

Db 606 HLYFYAQLKGLSR-----QKCPEEVKQMLHIIGLEDKWNSRSRF--LSGGMRRKLSIGI 657

Qy 207 QLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIH-QPRSELFQHFEDIA 265  
 |: ||:|||||:|:| :: | || : : || ::| | :| | :||

Db 658 ALIAGSKVLILDEPTSGMDAISRRRAIWDLL-QRQKSDRTIVLTTHFMDEADLLG--DRIA 714

Qy 266 ILTYGELVFCGTP---EEMLGFFNNGGYPC----PEHSNPF 300  
 |: ||| ||: :: | || | || |

Db 715 IMAGELQCCGSSLFLKQKYG----AGYHMTLVKEPHCNPED 752

RESULT 9

US-08-762-500-75

; Sequence 75, Application US/08762500

; Patent No. 6030806

; GENERAL INFORMATION:

; APPLICANT: Landes, Gregory M.

; APPLICANT: Burn, Timothy C.

; APPLICANT: Connors, Timothy D.

; APPLICANT: Dackowski, William R.

; APPLICANT: Van Raay, Terence J.

; APPLICANT: Klinger, Katherine W.

; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,

; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME

; NUMBER OF SEQUENCES: 83

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENZYME CORPORATION

; STREET: One Mountain Road

; CITY: Framingham

; STATE: Massachusetts

; COUNTRY: United States of America

; ZIP: 01701

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/762,500

; FILING DATE: 09-DEC-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/665,259

; FILING DATE: 17-JUN-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US96/10469

; FILING DATE: 17-JUN-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Dugan, Deborah A.

; REGISTRATION NUMBER: 37,315

; REFERENCE/DOCKET NUMBER: IG5-9.3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (508) 872-8400

; TELEFAX: (508) 872-5415

; INFORMATION FOR SEQ ID NO: 75:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1704 amino acids

```
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;      MOLECULE TYPE:  protein
US-08-762-500-75
```

Query Match 7.7%; Score 259.5; DB 3; Length 1704;  
Best Local Similarity 30.1%; Pred. No. 2.2e-16;  
Matches 85; Conservative 48; Mismatches 102; Indels 47; Gaps 13;

Qy	30	VTGTEARHSLGVLHVSYSVSNRVGPWWNIKSCQQKWDRQILKDVSLYIESGQIMCILGSS	89
		: :     :     :     : :     :     :     :	
Db	527	VAGIKIKH----LSKVFRVGNK-----DRAAVRDLNLNLNLYEQITVLLGHN	568
Qy	90	GSgKtTLlLDAISgRLRRTGTLEGEVfVNGCELRRD--QFQDCfSYVLQSDVfLSSlTVRE	147
		:       : :     : :     :     : :	
Db	569	GAGKTTTLsMLTGLfPPT---SGRAYISgYEISQDMVQIRKSLGLCPQHdILFDNLtVAE	625
Qy	148	TLRYtAML-ALCRSSADfYNKkVEAVMTELSlSHVADQMIGSYNfGGISSgERRRVsIAA	206
		:               : :   :   :         :       :	
Db	626	HLyFYAQlKGLSR-----QKCPeeVKQMLHIIGLEDKWNsRSRF--LSgGMRRKlSIGI	677
Qy	207	QLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVtIH-QPRSElFQHfDKIA	265
		:     :         :   :     : :     : :     :	
Db	678	ALIAGSKVLILDEPTSGMDAISRRAIWDLl-QRQKSDRTIVlTTHfMDEADlLG--DRIA	734
Qy	266	ILtYgELVfCGTP---EeMLGfFNnCGYPc---PEHSNPfD	300
		:         : :	
Db	735	IMAKgELQCCgSSlFLKQKYG---AGYHMTlVKEPHCNpED	772

## RESULT 10

US-09-543-681A-5411

; Sequence 5411, Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS  
MIRABILIS FOR

; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

```
; NUMBER OF SEQ ID NOS: 8344
```

; SEO ID NO 5411

; LENGTH: 653

```
; TYPE: PRT
```

; ORGANISM: Proteus mirabilis

US-09-543-681A-5411

Query Match 7.3%; Score 247; DB 4; Length 653;  
Best Local Similarity 21.8%; Pred. No. 8.1e-16;  
Matches 95; Conservative 89; Mismatches 136; Indels 116; Gaps 18;

Qy 66 DRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLLRRTGTLEGEVFNVG---CELR 122  
| :| :|| | :||:| :||:|||||:||||:| | | : : || | | ::

```

Db      26 DTVVLDQISLTINAGEMVAIIIGASGSGKSTLMN-ILGCLDKPSS--GEYKVAGQCVADME 82
Qy      123 RDQF----QDCFSYVLQSDVFLSSLTVRETLLRYTAMLALCRSSADFYNNKKVEAVMTELSL 178
      ||      :: | :: |      :: || : : | : | : : | :: | |
Db      83 SDQLAALRREHFGFIFQRYHLMHAHLTAEQNVEIPAIYA--GKSTEQRKERARALLTRLGL 140
Qy      179 SHVADQMIGSYNFGGIISSGERRRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLAE 238
      |::: | :| |:::| | | | | : : :|:: | | | : : : :| :
Db      141 ---AERI--HYRPSQLSGGQQQRVSIARALMNGGEVILADEPTGALDSQSGKEVMAILKQ 195
Qy      239 LARRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNGGYPCPEHSNP 298
      | :: | | : | | : | : | : | : | : | : | : | : |
Db      196 LNQQGHTVIVTHDPL--IAQQADRIIEIKDGQIISDN-----NN-----HHSAP 238
Qy      299 FDFYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLEPMVFFK 358
      | : : | : : :|:::
Db      239 -----VKKVPPAIQTASYFHQVI----- 256
Qy      359 TKDPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLMGLFLIFYLLRVQNNTLKGAVQ 418
      | : | | : | | : : : :|::: : : : : : | :
Db      257 -----GRFTQALNMAWRAMVVNKIRTLTLM-LGIIIGIASVVTII-----VIGDAAK 302
Qy      419 DRVGLLYQLVGATPYTGMLNAVNLFPMRLRAVSDQ-----ESQDGLYHKWQMLLAYV 469
      ||| : :| | : : :| | : | | : : :| :
Db      303 DRVLADIKAIKA-----NTIDIYPGKELGSDSPEDKQSLTIQDVKALKQQ-----SYI 350
Qy      470 LHVLPFSVIATVIFSS 485
      | | : |||
Db      351 QSVTP-----QIYFSS 361

```

RESULT 11

US-09-543-681A-8215

; Sequence 8215, Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS FOR

; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 8215

; LENGTH: 210

; TYPE: PRT

; ORGANISM: Proteus mirabilis

US-09-543-681A-8215

Query Match 7.3%; Score 246.5; DB 4; Length 210;

Best Local Similarity 33.7%; Pred. No. 1.4e-16;

Matches 69; Conservative 43; Mismatches 80; Indels 13; Gaps 4;

Qy 69 ILKDVSLYIESGQIMCILGSSGSGKTTLDAISGRLRRTGTLEGEVFNVCCELRRDQFD 128

```

Db      14 ILTEVSLHLEQGCCLGISGSSGSGKTTLLNAIAGYTDYTGDI---VLANQNMNKLFPVWQR 70
Qy      129 CFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTELSLSHVADQMIGS 188
      |: |      |||:: |      | : |      :| |: :: :: |:| :
Db      71 PCRYLNQRLYLFPFLTQNLWLAQYAAKQKRS----KEKEIALLEQMGIAHLATRYPSQ 126
Qy      189 YNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLLAELARRDRIVIV 248
      || ||::|||::| |: ||:::|||| : || | |: |: | :: :|:
Db      127 -----ISGGEQQRVALARALISQPKLLMDEPFSSLDWETRYQLWELIISLKKQQITMII 181
Qy      249 TIHQPRSELFQHFQDKIAILTYGELV 273
      |:|| ||      || :|: |:|
Db      182 VTHEPR-ELQALADKTLILLSNGKIV 205

```

RESULT 12

US-09-543-681A-7638

; Sequence 7638, Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS FOR

; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 7638

; LENGTH: 373

; TYPE: PRT

; ORGANISM: Proteus mirabilis

US-09-543-681A-7638

Query Match 7.3%; Score 246; DB 4; Length 373;

Best Local Similarity 26.4%; Pred. No. 4e-16;

Matches 74; Conservative 62; Mismatches 120; Indels 24; Gaps 7;

```

Qy      47 SVSNRVGPWWN----IKSCQKQWDR-QILKDVSLYIESGQIMCILGSSGSGKTTLLDAIS 101
      :||| |      | : :|| ::| ||:| : ||::| :|| ||||| ||| |:
Db      1 NVSNNAKQGQNMSEINHVTKYFDRTEVLHDVNLTVNSEMMALLGPSGSGKTTLLRIIA 60
Qy      102 GRLRRTGTLEGEVFNVCCELRRDQFQD-CFSYVLQSDVFLSSLTVRETLRY--TAMLALC 158
      | :| ||:: | :: | :: :| |      :|| | : : | :
Db      61 GLEHQT---EGKICFAGQDVSRLHARERKVGFFVQHYALFRHMTVFENIAFGLTVLPRRE 117
Qy      159 RSSADFYNNKKVEAVMTELSLSHVADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLD 218
      | : :||| :: : | |:| :      :| |::|||::| | :|:::|
Db      118 RPNKAAIDKKVTQLLEMIQLPHLAQRYPAQ-----LSGGQKQRVALARALAVEPQIILLD 172
Qy      219 EPTTGLDCMTANQIVLLLLAELARRDRIVIVVTIHQPRSELFQHFQDKIAILTYGELVFCGTP 278
      || ||      :: | || : | : | : | :| |: | :: |||
Db      173 EPFGALDAKVRTELRSWLRELHSELKFTSVFVTHDQOEAMEVADRIVIMGNKIEQVGT 232

```

QY 279 EEMLGFFNNCGYPCEHSNPFDFYMDLTSVDTQSREREIE 318  
:: : || : | : : ::  
Db 233 QQV-----WHTPESRFVLEFLGDVNHQLQGEINGAQLQ 264

## RESULT 13

```

US-09-134-001C-3832
; Sequence 3832, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3832
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3832

```

Query Match 7.2%; Score 243; DB 4; Length 242;  
Best Local Similarity 27.8%; Pred. No. 3.9e-16;  
Matches 74; Conservative 54; Mismatches 90; Indels 48; Gaps 12;

Qy 54 PWWNIKSCQQKWD-RQILKDVSLYIESGQIMCILGSSGSGKTT-----LLDAISGRLRR 106  
| |||: |: :||:|:| :| |:|: |:| ||||:| ||| |  
Db 2 PVINIKNLNKKFGANEVLRDINLTVEKGEVVAIIGPSGSGKSTLLRCMNLDDVPS----- 56

```

Qy      107 TGTLEGEVFEVNGCELRR-----DQFQDCFSYVLQS-DVFLSSLTVRETLRYTAMLALCRS 160
          :|:|      || :      | :      | |: ::|      :      :      ||| :
Db      57  ----KGKVI FDNELTQHNVHLDNLRQKMGMVFQNFNLFPHKKVIENVM--LAPLLHDKD 110

```

Qy 161 SADFYNNKKVEAVMTELSLSHVADQMIGSYNFGGIISSGERRRRVSIAAQLLQDPKVMMLDEP 220  
| | : | :: :: | || || : | |::|::| | : | |:: |||  
Db 111 SKDQLKEKALYLLEKVGGLKDKAD----SYP-NQLSGGQKQRVAIARALAMEPDVMLFDEP 165

```

Qy      221 TTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQHF-----DKIAILTYGELVFC 275
          | : | |          : : : : | |          : : | :          | | : | : |
Db      166 TSALDPEVVGDLKVMRQLANESMTMVIVTHE-----MNFakeISDKVVFEMADGVVVES 219

```

```

Qy      276  GTPEEMLGFFNNCGYPCPEHSNPFDF  301
          |||: :   | |       |:||   :|
Db      220  GTPQNI---FEN-----PQHSRTENF  237

```

RESULT 14

US-09-672-810-2  
; Sequence 2, Application US/09672810  
; Patent No. 6617450



```
; GENERAL INFORMATION:
; APPLICANT: STOCKER, PENNY J.
; APPLICANT: STEIMEL-CRESPI, DOROTHY T.
; APPLICANT: CRESPI, CHARLES L.
; TITLE OF INVENTION: P-GLYCOPROTEINS AND USES THEREOF
; FILE REFERENCE: G0307/7018
; CURRENT APPLICATION NUMBER: US/09/672,810
; CURRENT FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/156,921
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/158,818
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1280
; TYPE: PRT
; ORGANISM: Macaca fascicularis
US-09-672-810-2
```

```
Query Match          7.2%; Score 242; DB 4; Length 1280;
Best Local Similarity 21.9%; Pred. No. 8.1e-15;
Matches 125; Conservative 92; Mismatches 181; Indels 174; Gaps 24;
```

```
Qy      15 PHINRGSLSSELEQGSVTGT-EARHSLGVLHVSYSVSNRVGPWWNIKSCQQKWDRQILKDV 73
      | | : | | : : | | | : | | | : | | | :
Db      373 PSIDSYSKSGHKPDNIKGNLEFRN----VHFSYPSRKEV-----KILKGL 413

Qy      74 SLYIESGQIMCILGSSSGSKTTLLDAISGRLLRRTGTLEGEVFNVCCELRRDQ---FQDCF 130
      : | : | | | : : | | | | | : : | | | | : | : | :
Db      414 NLKVQSGQTVALVGNSGCGKSTTVQLMQ---RLYDPTGEMVSDGQDIRTINVRFLREII 470

Qy     131 SYVLQSDVFLSSLTVRETLYRTAMLALCRSSADFYNNKKVEAVMTE-----LSLSHVAD 183
      | | | | : | : | : | | | | : | : | : | : |
Db     471 GVVSQEPV-LFATTIAENIRY-----GREDVTMDEIEKAVKEANAYDFIMKLPQKFD 521

Qy     184 QMIGSYNFGGSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLIA-ELARR 242
      : : | : | : | : | : | : | : | : | : | : | : | :
Db     522 TLVGERG-AQLSGGQKQRIAIARALVRNPKILLLDEATSALD--TESEAVVQVALDKARK 578

Qy     243 DRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCGYPCPEHSNPFDFY 302
      | | | | | : | | | | : | | | : | : |
Db     579 GRTTIVIAH--RLSTVRNADVIAGFDDGVIVEKGNHDELM-----KEKGIY 622

Qy     303 MDLTSVDTQSREREI-----ETYKRVQMLECAFKES-----DIYHKILENIERARYL 349
      | : : | | | | : | : | : | : | : | : | : |
Db     623 FKLVTMQTAGNEIELENAADESKSEIDTLEMSSHDSGSSLIRKRSTRRSVRGSQGQDRKL 682

Qy     350 KT-----LPMVPF-----KTKDP-----PGMFGKLGVLRLRVTRNLMRN 383
      | : | | | | : | | | : | : | : | : |
Db     683 STKEALDESIPPVSWRIMKLNLTWPYFVVGVFCAIINGGLQPAFAVIFSKIIGIFTRN 742

Qy     384 KQAVIMRLVQNLMGLFLI-----FYL----- 405
      | | | | | : | : | : | : |
Db     743 DDAETKRQNSNLFSLFLVLGIVSFITFFLQGFTEGKAGEILTKRLRYMVFRSMLRQDVS 802

Qy     406 -----LRVQNNT--LKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMRAVS 450
```

```

      | : | : : | | : | : : | : | | : : :
Db      803 WFDDPKNTTGALTTRLANDAAQVKGAIGSRLAIITQNI-ANLGTGIIIS----- 850

Qy      451 DQESQDGLYHKWQMLLAYVLHVLPPFSVIATVI 482
      | : | | : | : | : | | | :
Db      851 -----LIYGWQLTL-LLLAIVPIIAIAGVV 874

```

RESULT 15

US-09-672-810-4

```

; Sequence 4, Application US/09672810
; Patent No. 6617450
; GENERAL INFORMATION:
; APPLICANT: STOCKER, PENNY J.
; APPLICANT: STEIMEL-CRESPI, DOROTHY T.
; APPLICANT: CRESPI, CHARLES L.
; TITLE OF INVENTION: P-GLYCOPROTEINS AND USES THEREOF
; FILE REFERENCE: G0307/7018
; CURRENT APPLICATION NUMBER: US/09/672,810
; CURRENT FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/156,921
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/158,818
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1283
; TYPE: PRT
; ORGANISM: Macaca fascicularis
US-09-672-810-4

```

```

Query Match          7.2%; Score 242; DB 4; Length 1283;
Best Local Similarity 21.9%; Pred. No. 8.1e-15;
Matches 125; Conservative 92; Mismatches 181; Indels 174; Gaps 24;

```

```

Qy      15 PHINRGSLSSLEQGSVTGT-EARHSLGVLHVSYSVSNRVGPWWNIKSCQQKWDRQILKDV 73
      | | : | | : : | | | : : | | | : | | :
Db      376 PSIDSYSKSGHKPDNIKGNLEFRN----VHFSYPSRKEV-----KILKGL 416

Qy      74 SLYIESGQIMCILGSSSGSKTTLDDAISGRLRRTGTLEGEVFNVCCELRRDQ---FQDCF 130
      : | : : | | : : | : | | | : : | | | : | : :
Db      417 NLKVQSGQTVALVGNSGCGKSTTVQLMQ---RLYDPTEGMVSDGQDIRTINVRLREII 473

Qy      131 SYVLQSDVFLSSLTVRETLYRTAMLALCRSSADFYNNKKVEAVMTE-----LSLSHVAD 183
      | | | | : | : | : | | | : : | | : | : | :
Db      474 GVVSQEPV-LFATTIAENIRY-----GREDVTMDEIEKAVKEANAYDFIMKLPQKFD 524

Qy      184 QMIGSYNFGGISSGERRRVISIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLA-ELARR 242
      : : | : | : | : | : | : | : | : | : | : | : | :
Db      525 TLVGERG-AQLSGGQKQRIAIARALVRNPKILLLDEATSALD--TESEAVVQVALDKARK 581

Qy      243 DRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPFDFY 302
      | | | | | : | | | : | | | : : : |
Db      582 GRTTIVIAH--RLSTVRNADVIAGFDDGVIVEKGNHDELM-----KEKGIY 625

Qy      303 MDLTSVDTQSREREI-----ETYKRVQMLECAFKES-----DIYHKILENIERARYL 349

```

	: :     :   : :   : :	
Db	626 FKLVTMQTAGNEIELENAADESKSEIDTLEMSSHDSGSSLIRKRSTRRSVRSQGQDRKL	685
Qy	350 KT-----LPMVPF-----KTKDP-----PGMFGKLGVLRLRVTRNLMRN	383
	:       :   : :	
Db	686 STKEALDESIPPVSFWRIMKLNLEWPFVVGVFCAIINGGLQPAFAVIFSKIIGIFTRN	745
Qy	384 KQAVIMRLVQNLIMGLFLI-----FYL-----	405
	:   :	
Db	746 DDAETKRQNSNLFSLFLVLGIVSFITFFLQGFTEGKAGEILTKRLRYMVFRSMLRQDVS	805
Qy	406 -----LRVQNT--LKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMRLRAVS	450
	:   : :   :   :   :	
Db	806 WFDDPKNTTGALTTRLANDAAQVKAIGSRLAITQNI-ANLGTGIIIS-----	853
Qy	451 DQESQDGLYHKWQMLLAYVLHVLFPFSVIATVI	482
	:   :   :   :	
Db	854 -----LIYGWQLTL-LLLAIVPIIAIAGVV	877

Search completed: February 27, 2004, 07:20:12  
Job time : 16.7734 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2004, 06:44:33 ; Search time 14.5272 Seconds  
(without alignments)  
4317.206 Million cell updates/sec

Title: US-09-989-981A-2  
Perfect score: 3369  
Sequence: 1 MGELPFLSPEGARGPHINRG.....PALVILGIVIFKVRDYLSR 652

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pirl:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	685	20.3	656	2	JC7860	brain multidrug re
2	630	18.7	725	2	C84423	probable ABC trans
3	628	18.6	1294	2	S77690	probable membrane
4	606.5	18.0	725	2	T47652	ABC transporter-li
5	602.5	17.9	1049	1	S19421	ATP-dependent perm
6	597	17.7	590	2	B96573	protein F12M16.17
7	595	17.7	609	2	E96742	probable ABC trans
8	589	17.5	687	1	FYFFW	white protein - fr
9	587	17.4	708	2	T47650	ABC transporter-li
10	580	17.2	638	2	G02068	white homolog - hu
11	579.5	17.2	658	2	T31958	hypothetical prote
12	577	17.1	720	2	T47648	ABC transporter-li
13	573.5	17.0	608	2	T34391	hypothetical prote

14	571.5	17.0	687	2	D96553	hypothetical prote
15	570	16.9	559	2	B88474	protein C05D10.3 [
16	569.5	16.9	646	2	JC7777	ATP binding casset
17	568.5	16.9	646	2	C86441	probable ABC trans
18	562	16.7	755	2	G84791	probable ABC trans
19	559	16.6	740	1	T02567	probable ATP-bindi
20	556	16.5	649	2	A84509	probable ABC trans
21	554	16.4	610	2	T19333	hypothetical prote
22	545	16.2	635	2	T08934	hypothetical prote
23	542	16.1	739	2	T45891	ABC transporter-li
24	539	16.0	662	2	T47649	ABC transporter-li
25	536.5	15.9	633	2	T19189	hypothetical prote
26	533.5	15.8	577	2	T04229	ABC-type transport
27	523.5	15.5	639	2	G88839	protein C10C6.5 [i
28	521.5	15.5	659	2	E86313	hypothetical prote
29	520	15.4	695	2	T21109	hypothetical prote
30	515	15.3	547	2	T31543	hypothetical prote
31	512.5	15.2	678	2	H96552	hypothetical prote
32	487	14.5	737	2	T46101	ABC transporter-li
33	473.5	14.1	705	2	D84680	probable ABC trans
34	471	14.0	1450	2	A84780	probable ABC trans
35	465.5	13.8	675	1	FYFFB	brown protein - fr
36	457	13.6	1333	2	S63403	probable membrane
37	454.5	13.5	668	2	S55023	brown protein - fr
38	450	13.4	1450	2	T45888	ABC transporter-li
39	442.5	13.1	1426	2	T30567	ATP-binding casset
40	431.5	12.8	1435	2	D96693	protein Putative A
41	431.5	12.8	1443	2	T02491	probable ABC trans
42	429	12.7	1451	2	B86286	F9L1.15 protein -
43	428.5	12.7	1469	2	H96622	probable ABC trans
44	423.5	12.6	1413	2	G84790	probable ABC trans
45	421.5	12.5	1420	2	T02644	ABC-type transport

#### ALIGNMENTS

##### RESULT 1

JC7860

brain multidrug resistance protein, BMDP - pig

C;Species: *Sus scrofa domestica* (domestic pig)

C;Date: 18-Nov-2002 #sequence\_revision 18-Nov-2002 #text\_change 31-Mar-2003

C;Accession: JC7860

R;Eisenblaetter, T.; Galla, H.J.

Biochem. Biophys. Res. Commun. 293, 1273-1278, 2002

A;Title: A new multidrug resistance protein at the blood-brain barrier.

A;Reference number: JC7860; MUID:22050127; PMID:12054514

A;Accession: JC7860

A;Molecule type: mRNA

A;Residues: 1-656 <EIS>

A;Cross-references: GB:AJ420927

A;Experimental source: brain

C;Comment: This protein, a new transport protein of the ATP-binding cassette (ABC) superfamily of transporters, expressed in porcine brain capillary endothelial cells, plays an importnat role in the exclusion of xenobiotics from the brain and participates in drug transport across the blood-brain barrier and therefore is considered as a efflux4 pump at the cerebral endothelium.

C;Genetics:  
A;Gene: bmdp

Query Match 20.3%; Score 685; DB 2; Length 656;  
Best Local Similarity 29.7%; Pred. No. 3.8e-43;  
Matches 187; Conservative 130; Mismatches 228; Indels 84; Gaps 19;

```
Qy      31 TGTEARHSLGLVHVSY-SVSNRVGPWWNIKS----CQQKWDRQILKDVSILYIESGQIMCI 85
      : | : | | : | : | | : | | : | : | : | : | : | : | : |
Db      24 SSNELKTSAGGAVLSFHDICYRV----KVKSGFLFCRKTVEKEILTNINGIMKPG-LNAI 78

Qy      86 LGSSSGSGKTTLLDAISGRLRRTGTLEGEVFNVCCELRRDQFQDCFSYVLQSDVFLSSLTV 145
      || : | || : || | : | | | : | | | : | | | : | | | : | | |
Db      79 LGPTGGGKSSLLDVLAARKDPHG-LSGDVLINGAP-RPANFKCNSGYVVDVVMGTLTV 136

Qy     146 RETLRYTAMLALCRSSADF-YNKKVEAVMTELSLSHVADQMIGSYNFGGSISSGERRRVSI 204
      || | : : | | | : : | : : | : | | | | | : | : | | | | |
Db     137 RENLQFSAALRLPTTMTNHEKNERINMVIQELGLDKVADSKVGTQFIRGVSGGERKRTSI 196

Qy     205 AAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQHFDDKI 264
      | : | : | | : | | | | | | | | | | : | : | | | | : | : | :
Db     197 AMELITDPSILFLDEPTTGLDSSSTANAVLLLLKRMASKQGRTIIFSIIHQPRYSIFKLFDSL 256

Qy     265 AILTYGELVFCGTPPEMLGFFNCCGYPCPEHSNPFDFYMDLTSVDTQ-----S 312
      : | | | : | | | | | : | | | : | | | | : | | : | : :
Db     257 TLLASGRLMFHPAREALGYFASIGYNCEPYNPNADFFLDVINGDSSAVVLSRADRDEGA 316

Qy     313 REREIETYKRVQMLE--CAF-----KESDIYHKILENIERAR 347
      : | | | | : : : | | | : | | | | : | : | : | :
Db     317 QEPEEPPEKDTPLIDKLAIFYTNSSFFKDTKVELDQFSGGRKKKKSSVYKEVYTTTSFCH 376

Qy     348 YLKTLPMPVFPKTKDPPGFMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFL--IFYL 405
      | : : | : : | | : | : | | : : : | : | : | |
Db     377 QLRWIS-----RRSFKNLLGNPQASVAQIIVTIILGLVIGAIIFYD 416

Qy     406 LRVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQML 465
      | : | | | : | : | : | : : : | : : | | |
Db     417 LK---NDPSG-IQNRAVGLFFLTNQCFS--VSAVELLVVEKKLFIHEYISGYRVSSYF 471

Qy     466 LAYVL-HVLPFSVIATVIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFLLVLLGI 524
      : | : | | : : : | : | | | | | | : : : : : | |
Db     472 FGKLLSDLLPMRMLPSIIFTCITYFLLGLKPAVGSGFFIMMFTLM---MVAYSASSMALAI 528

Qy     525 VQNPNIIVNSIVALLSIS--GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCILVVNEFY 582
      : : | : | : | | : : | | | : : : | | | : | | |
Db     529 AAGQSVSVATLLMTISFVFMIFSGLLVNLKTVVPWLSWLQYFSIPRYGFSALQYNEFL 588

Qy     583 GLNFTCGGSNTSMLNHPMCAITQGVQFIE 611
      | | | | | : | | | : : |
Db     589 GQNF-CPGLNVTNNTCSFAICTGAEYLE 616
```

## RESULT 2

C84423

probable ABC transporter [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C;Accession: C84423

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell, C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams, M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhagen, G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.C.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: C84423

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-725 <STO>

A;Cross-references: GB:AE002093; NID:g4262239; PIDN:AAD14532.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g01320

A;Map position: 2

Query Match 18.7%; Score 630; DB 2; Length 725;  
Best Local Similarity 30.2%; Pred. No. 5.6e-39;  
Matches 188; Conservative 104; Mismatches 244; Indels 86; Gaps 18;

```
Qy      55 WWNKSC-----QKQWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISG-----R 103
      | || :|      | | :||:| | :|::| |:| ||||| |:| :| |
Db      72 WRNI-TCSLSDKSSKSVRFLLNKVSGEAKPGRLLAIMGPSGSGKTTLLNVLAGQLSLSPR 130

Qy      104 LRRTGTLEGEVFNVCCELRRDQFQDCFSYVLQSDVFLSSLTVRETLYTAMLALCR-SSA 162
      | :| ||      |||      : : :| | | :| | ||||| : | | | |||
Db      131 LHLSGLLE----VNGKPSSSKAYK--LAFVRQEDLFFSQLTVRETLSFAAELQLPEISSA 184

Qy      163 DFYNKKVEAVMTELSLHVADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTT 222
      : : :| : : :| | || :|      ||| ||::|:|:| :|: | |: ||||
Db      185 EERDEYVNNLLKLGLVSCADSCVGDQKVRGISGGEKKRLSLACELIASPSVIFADEPTT 244

Qy      223 GLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCG-TPEEM 281
      ||| | : : :| :||:      || :||| | : : || | :|| | ||: | :|
Db      245 GLDAFQAQKVMETLQKLAQDGHTVICSIHQPRGSVYAKFDDIVLLTEGTLVYAGPAGKEP 304

Qy      282 LGFFNNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILE 341
      | :| | | : |||| || :| || ||| | | : ||| | || :
Db      305 LTYFGNFGFLCPEHVNPAEFLADLISVDYSSSETVYSSQKRVHALVDAFSQ----- 355

Qy      342 NIERARYLKTLPMPVFKTKD-----PPGMFGKLGVLRLR----VTRNLM 381
      |      |      | :      | : : :||:|      :|:
Db      356 -----RSSSVLYATPLSMKEETKNMRRPRKAIVERTDGWWRQFFLLLKRAWMQASRDGP 410

Qy      382 RNKQAVIMRLVQNLIMGLFLIFYLLRVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVN 441
      ||      | : :| | :|: :      : : |||:| |      | : |
Db      411 TNKVRARMSVASAVIFG--SVFWRM----GKSQTSIQDRMGLLQVAINTAMAALTKTVG 464

Qy      442 LFPMLRAVSDQESQDGLYHKWQMLLAYVLHVLPFSVIATVIFSSVCYWTGLGYPEVARFG 501
      :|| ||: |:| | | ||: : :|      :| :| | | :|||
Db      465 VFPKERAIVDRERSKGSYSGLPYLLSKTIAEIPIGAAFPLMFGAVLYPMARLNPTLSRFG 524
```

```

Qy      502 YFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSI-VALLSISGLLIGSGFIRNIQEMPIPL 560
      | : : | : | : : : : : : : | : | ||
Db      525 KFCGIVTVESFAASAMGLTVGAMVPSTEAAVGPSTLMTV--FIVFGGYVNADNTPIIF 582

Qy      561 KILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPMCAITQGVQFIEKTCPG----- 616
      : : : : : | : || || | || : | | : | : |
Db      583 RWIPRASLIRWAFQGLCINEFSGGLKF--DHQNT-----FDVQTGEQALERLSFGGRRI 633

Qy      617 -----ATSR-----FTANFLIL 628
      | || : | : | : |
Db      634 RETIAAQRILMFWYSATYLLL 655

```

### RESULT 3

S77690

probable membrane protein YOL075c - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: hypothetical protein O1125; hypothetical protein O1130; hypothetical protein YOL074c

C;Species: *Saccharomyces cerevisiae*

C;Date: 21-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 19-Apr-2002

C;Accession: S77690; S66767; S66768

R;Alexandraki, D.; Katsoulou, C.; Tzermia, M.

submitted to the Protein Sequence Database, July 1996

A;Reference number: S66756

A;Accession: S77690

A;Molecule type: DNA

A;Residues: 1-1294 <ALE>

A;Cross-references: EMBL:Z74816; MIPS:YOL075c

A;Note: this is a revision to the sequence from reference S66756

A;Accession: S66767

A;Molecule type: DNA

A;Residues: 1-179, 'TTRTGVFLVVKRED' <ALW>

A;Cross-references: EMBL:Z74816

A;Experimental source: strain S288C

A;Note: this sequence has been revised in reference S77690

A;Note: this was assumed to be protein YOL074c

A;Accession: S66768

A;Molecule type: DNA

A;Residues: 200-1294 <ALF>

A;Cross-references: EMBL:Z74817

A;Experimental source: strain S288C

A;Note: this sequence has been revised in reference S77690

A;Note: this was assumed to be the complete sequence of protein YOL075c

C;Genetics:

A;Cross-references: SGD:S0005435

A;Map position: 15L

A;Note: YOL075c

C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

C;Keywords: ATP; nucleotide binding; P-loop; transmembrane protein

F;45-263/Domain: ATP-binding cassette homology <ABC1>

F;62-69/Region: nucleotide-binding motif A (P-loop)

F;376-392/Domain: transmembrane #status predicted <TM1>

F;469-485/Domain: transmembrane #status predicted <TM2>

F;496-512/Domain: transmembrane #status predicted <TM3>

F;606-622/Domain: transmembrane #status predicted <TM4>

F;710-916/Domain: ATP-binding cassette homology <ABC2>



F;727-734/Region: nucleotide-binding motif A (P-loop)  
 F;1042-1058/Domain: transmembrane #status predicted <TM5>  
 F;1125-1141/Domain: transmembrane #status predicted <TM6>  
 F;1177-1193/Domain: transmembrane #status predicted <TM7>  
 F;1269-1285/Domain: transmembrane #status predicted <TM8>

Query Match 18.6%; Score 628; DB 2; Length 1294;  
 Best Local Similarity 29.6%; Pred. No. 1.7e-38;  
 Matches 183; Conservative 123; Mismatches 237; Indels 76; Gaps 21;

```

Qy      67 RQILKDVS LYIESGQIMC ILGSSGSGKT TLLDAISGR LRRT----GTLEGEVFVNGCEL R 122
      ::||: | : | | | :| ||||: ||: ||||: : | : | ::
Db      707 KEILQSVNAIFKPGMINAIMGPSGSGKSSLLNLISGR LKSSVFAKFDTS GSIMFNDIQVS 766

Qy      123 RDQFQDCFSYVLQ-SDVFLSSLT VRET LRYTAMLALCRSSADFY NKKVEAVMT ELSLHV 181
      |:: ||| | | |::|||: |||: | | | : : : : : | | |
Db      767 ELMFKNVCSYVSQDDDHL LAALTVKETLKYAAALRLHHLTEAERMERTDNLIRSLGLKHC 826

Qy      182 ADQMIGSYNFGGISSGERRRVSIAAQL LQDPKVMMLDEPTTGLDCMTANQIV LLLAE LAR 241
      : :||: ||| ||: ||: : ||| || : :||| ||: || | : | : | |
Db      827 ENNIIGNEFVKGISGGEKRRVTMGVQL LNDPPILLLDEPTSG LDSFTSATILEILEKLCR 886

Qy      242 -RDRIVIVTIHQPRSELFQHFDKIAILT-YGELVFCGTP EEMLGFFNNCGYPCPEHSNPF 299
      : : :|: ||||| |||: | : :| | | | :|: ||: :| || || :|
Db      887 EQGKTIIITIHQPRSELFKRFGNVLL LAKSGRTAFNGSPDEMIAYFTELGYNCPSFTNVA 946

Qy      300 DFYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILEN-----IERARYLKT 351
      ||: || ||: ||: : || : ||: : | :| : :| | :| :
Db      947 DFFDLISVNTQNEQNEISSRARVEKILSAWKA N-----MDNESLSPTPISEKQQYSQE 1000

Qy      352 LPMVVPFK--TKDPPGMFGKLGVL LRRVTRNLMRNKQAVIMRLVQN LIMGLFLIFYLLRVQ 409
      : : | : | :| | :| : : :| :| :| :| :| :| :
Db      1001 SFFTEYSEFVRKPANLVLAYIVNVKRQFTT TRRSFDSL MARIAQIPGLGVIFALFFAPVK 1060

Qy      410 NNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYV 469
      :| : : :|: || : | : ||| : :| | :| | :| |||:
Db      1061 HNYT--SISNRLGLAQEST-ALYFVGMLGNLACYPTERDYFYEEYNDNVYGIAPFFLAYM 1117

Qy      470 LHVLPFSVIATVIFSSVCYWT LGLYPEVARFGYFSAALLAPHLI---GEFLT LVLLGIVQ 526
      || | :|: ||: || | | | | : : : || | : :
Db      1118 TLELPLSALASVLYAVFTVLACGL-PRTA--GNFFATVYCSFIVTCCGERLGIMTNTFFE 1174

Qy      527 NPN-IVNSIVALLSI----SGLL-IGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVNE 580
      | :|| | :||| |||: :| | || | | | :| :| :
Db      1175 RPFQVNCISIIISIGTQMSGLMSLG-----MSRVLKGFNYLNPVGYSMIINFA 1225

Qy      581 FYG-LNFTC--GGSNTSMLNHPMCAITQGVQFIEKTCPGATS RFTANFLILYGFIP---- 633
      | | | || || | : : || | :| :| :| :
Db      1226 FPGNLKLTCEDGGKNS-----DGTCEFANGH---DVLVSYGLVRNTQK 1265

Qy      634 --ALVILGIVIFKVRDYLI 650
      : : : :|: : : |
Db      1266 YLGIIVCVAIIRLIAFFI 1284

```

ABC transporter-like protein - Arabidopsis thaliana  
 N;Alternate names: protein T26I12.10  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 19-May-2000  
 C;Accession: T47652  
 R;Monfort, A.; Casacuberta, E.; Puigdomenech, P.; Mewes, H.W.; Lemcke, K.;  
 Mayer, K.F.X.; Quetier, F.; Salanoubat, M.  
 submitted to the Protein Sequence Database, February 2000  
 A;Reference number: Z24471  
 A;Accession: T47652  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-725 <MON>  
 A;Cross-references: EMBL:AL132954  
 A;Experimental source: cultivar Columbia; BAC clone T26I12  
 C;Genetics:  
 A;Map position: 3  
 A;Note: T26I12.10  
 C;Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein  
 F12L6.1; ATP-binding cassette homology

Query Match 18.0%; Score 606.5; DB 2; Length 725;  
 Best Local Similarity 28.9%; Pred. No. 3.2e-37;  
 Matches 173; Conservative 122; Mismatches 226; Indels 77; Gaps 15;

```

Qy      62 QQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVCCEL 121
      :|  : :| |||  | | : ||: ||: ||: ||: ||: ||:  | :| | | : || :
Db      92 RQNGVKTLDDVSGEASDGDILAVLGASGAGKSTLIDALAGRAE-GSLRGSVTLNGEKV 150

Qy     122 RRDQFQDCFS-YVLQSDVFLSSLTVRETLYRTAMLALCRS-SADFYNNKKVEAVMTLSLS 179
      : :  | ||: | | : ||: ||| : :  | || |  : |||: : | |
Db     151 LQSRLKVISAYVMQDDLLFPMLTVKETLMFASEFRLPRSLSKSKKMERVEALIDQLGLR 210

Qy     180 HVADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAEL 239
      : | : :||  | :| ||||| ||  : : || | : ||||: |||  | :| :| :
Db     211 NAANTVIGDEGHRGVSGGERRRVSIGIDIIHDPIVLFLDEPTSGLDSTNAFMVQVLKRI 270

Qy     240 ARRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPPEMLGFFNNCGYPCPEHSNPF 299
      | :  |||: ||| | : : : | : : ||: | : || | : | ||: : | | || |
Db     271 AQSGSIVIMSIHQPSARIVELLDRLLIILSRGKSVFNGSPASLPGFFSDFGRPIPEKENIS 330

Qy     300 DFYMDLTSVDTQSREREIE-----TYKRVQMLECAF 330
      :| :||  ||:|  | | |  :
Db     331 EFALDLV-----RELEGSNEGTKALVDFNEKWQONKISLIQSAPQTNKLDQDRSLSL 382

Qy     331 KESDIYHKILENIERARYL-----KTLPMVPFKTKDPPGMFGKLGVLLRRVTRNLMRNK 384
      ||:  |  : : | : :  |  :  | : | : :| :| :| :|
Db     383 KEA-----INASVSRGKLVSGSSRSNPTSMETVSSYANPSLF-ETFILAKRYMKNWIRMP 436

Qy     385 QAVIMRLVQNLIMGLFL--IFYLLRVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNL 442
      : | | :  : : | | : : |  : :| : || | :| : | : | | : | :
Db     437 ELVGTRIATVMVTGCLLATVYWKL---DHTPRGA-QERL-TLFAFVVPTMFYCCLDNVPV 491

Qy     443 FPMLRAVSDQESQDGLYHKWQMLLAYVLHVLPFSVIATVIFSSVCYWTGLYPEVARFGY 502
      | | : :| :  |  : : : | || :  : : ||: : ||: ||  : | :
Db     492 FIQERYIFLRETTNAYRTSSYVISHSLVSLPQLLAPSLVFSAITFWTVGLSGGLEGFVF 551

```

```

Qy      503 FSAALLAPHLIGEFLLTLVLLGIVQNPNIIVNS-IVALLSISGLLIGSGFIRNIQEMPIPLK 561
      :   : |   | :   : | : | | | :   : | : :   : : | : | | | |   : |
Db      552 YCLLIYASFWSGSSVVTFIGSVV--PNIMLCYMVSTITYLAYCLLLSGFYVNRDRIPFYWT 609

Qy      562 ILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPMCAITQGVQFIEKTCPGATS 619
      | : | | | : : : | | | : |   : | | | : | | | |
Db      610 WFHYISILKYPYEAVLINEF-----DDPSRCFVRGVQVFDSTLLGGVS 652

```

# RESULT 5

S19421

ATP-dependent permease ADP1 precursor - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: protein YCR011c; protein YCR105

C;Species: *Saccharomyces cerevisiae*

C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 19-Jan-2001

C;Accession: S19421; S40914

R;Goffeau, A.; Purnelle, B.; Skala, J.

submitted to the Protein Sequence Database, March 1992

A;Reference number: S19420

A;Accession: S19421

A;Molecule type: DNA

A;Residues: 1-1049 <GOF>

A;Cross-references: EMBL:X59720; NID:g1907116; PIDN:CAA42328.1; PID:g1907154;

GSPDB:GN00003; MIPS:YCR011c

R;Purnelle, B.; Skala, J.; Goffeau, A.

Yeast 7, 867-872, 1991

A;Title: The product of the YCR105 gene located on the chromosome III from *Saccharomyces cerevisiae* presents homologies to ATP-dependent permeases.

A;Reference number: S40914; MUID:92160395; PMID:1789009

A;Accession: S40914

A;Status: not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 1-1049 <PUR>

R;Skala, J.; Purnelle, B.; Goffeau, A.

Yeast 8, 409-417, 1992

A;Title: The complete sequence of a 10.8 kb segment distal of SUF2 on the right arm of chromosome III from *Saccharomyces cerevisiae* reveals seven open reading frames including the RVS161, ADP1 and PGK genes.

A;Reference number: S25353; MUID:92327849; PMID:1626432

A;Contents: annotation

C;Genetics:

A;Gene: SGD:ADP1; MIPS:YCR011c

A;Cross-references: SGD:S0000604; MIPS:YCR011c

A;Map position: 3R

C;Superfamily: ATP-dependent permease ADP1; ATP-binding cassette homology

C;Keywords: ATP; glycoprotein; nucleotide binding; P-loop; transmembrane protein

F;1-25/Domain: signal sequence #status predicted <SIG>

F;26-1049/Product: ATP-dependent permease ADP1 #status predicted <MAT>

F;26-324/Domain: extracellular #status predicted <EXT>

F;325-341/Domain: transmembrane #status predicted <TM1>

F;406-607/Domain: ATP-binding cassette homology <ABC>

F;423-430/Region: nucleotide-binding motif A (P-loop)

F;550-557/Region: nucleotide-binding motif B

F;794-810/Domain: transmembrane #status predicted <TM2>

F;829-845/Domain: transmembrane #status predicted <TM3>

F;878-894/Domain: transmembrane #status predicted <TM4>

F;909-925/Domain: transmembrane #status predicted <TM5>

F;938-954/Domain: transmembrane #status predicted <TM6>  
 F;1025-1041/Domain: transmembrane #status predicted <TM7>  
 F;50,114,165,221/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F;429/Binding site: ATP (Lys) #status predicted

Query Match 17.9%; Score 602.5; DB 1; Length 1049;  
 Best Local Similarity 26.5%; Pred. No. 1e-36;  
 Matches 191; Conservative 130; Mismatches 227; Indels 173; Gaps 25;

Qy	38	SLGVLHVSYSVSNRVGPWWNIKSCQQKWDRQILKDVSLYIESGQIMCILGSSSGSKTTLL	97
		:  :::        :: :  :  :  :    :  :  :  :	
Db	383	TLSFENITYSV-----PSINSDGVEE-----TVLNEISGIVKPGQILAIMGGSGAGKTTLL	433
Qy	98	DAISGRLRRTGTLEGEVFNVCCELRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLAL	157
		:: :  ::  :   :     :     :        :      : : :	
Db	434	DILAMK-RKTGHVSGSIKVNIGISMDRKSFSKIIGFVDQDDFLPTLTVFETVLNSALLRL	492
Qy	158	CRS-SADFYNNKKVEAVMTELSLSHVADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVM	216
		::   : :    :    : :  ::  :       :      : :    :	
Db	493	PKALSFEAKKARVYKVLEELRIIDIKDRIIGNEFDRGISGGEKRRVSIACELVTSPLVLF	552
Qy	217	LDEPTTGLDCMTANQIVLLLLAELAR-RDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFC	275
		:        ::     : :  ::::       :      : :  :: :	
Db	553	LDEPTSGLDASNANNVIECLVRLSSDYNRTLVLSTHQPRSNIFYLFDKLVLLSKGEMVYS	612
Qy	276	GTPEEMLGFFNCCGYPCPEHSNPFDFYMDLT-----SVD-	309
		::           ::     : : :   :	
Db	613	GNAKKVSEFLRNEGYICPDNYNIADYLIDITFEAGPQGKRRRIRNISDLEAGTDTNDIDN	672
Qy	310	-----TQSREREIETYKRVQMLECA	329
		:	
Db	673	TIHQTTFTSSDGTQREWAHLAAHRDEIRSLRDEEDVEGTDGRRGATEIDLNTKLLHDK	732
Qy	330	FKESDIYHKILENI-----ERARYLK-TLPMVPFKTKDPPGMFGKLGVLRLRVTRNL	380
		: :    :: :     :      :   :  :    : :	
Db	733	YKDSVYYAELSQEIEEVLSEGDEESNVLNGDLP----TGQQSAGFLQQLSILNSRSFKNM	788
Qy	381	MRNKQAVIMRLVQNLMGLFL--IFYLLRVQNNTLKGAVQDRVGLLYQLV---GATPYTG	435
		: :: : ::    ::  : :  :    ::   : ::   :	
Db	789	YRNPKLLLGNYYLLTILLSLFLGTLYYNV---SNDISG-FQNRMGLEFFILTIFYGFVTFGTG	844
Qy	436	MLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYVL-----HVLPSVIATVIFSSVCYWT	490
		: :   :   : :  :       :   :    : :  :	
Db	845	L----SSFALERIIFIKERSNNYYSP----LAYYISKIMSEVVPLRVVPPILLSLIVYPM	896
Qy	491	LGLYPEVARFGYFSAALLAPHLIGEFLTLVLVLGIV---QNPNIIVNSIVALLS---ISGLL	544
		:   : :  :     : :  :   :  : :	
Db	897	TGLNMKDNAF-FKCIGILILFNLGISLEILTIGIIFEDLNNSIILSVLVLLGSLLFSGLF	955
Qy	545	IGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVNEF-----YGLNFTCGGSNTSM	595
		:     :       :      ::	
Db	956	INTKNITN-----VAFKYLKNFSVFYYAYESLLINEVKTIMLKERKYGLNI-----	1001
Qy	596	LNHPMCAITQGVQFIEKTCPGATSRTANFLILYGFIPALVILGI-----VIFKVRDY	648
		:  : ::     :  :	
Db	1002	-----EVPGAT-----ILSTFGFVVQNLVFDIKILALENVVFLIMGY	1038

Qy 649 L 649  
|  
Db 1039 L 1039

RESULT 6

B96573

protein F12M16.17 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001

C;Accession: B96573

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;  
Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.;  
Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.;  
Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.;  
Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.;  
Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.;  
Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-  
Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu,  
S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzialli, A.; Militscher, J.; Miranda,  
M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.;  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;  
Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken,  
S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.;  
Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: B96573

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-590 <STO>

A;Cross-references: GB:AE005173; NID:g7769856; PIDN:AAF69534.1; GSPDB:GN00141

C;Genetics:

A;Gene: F12M16.17

A;Map position: 1

C;Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 17.7%; Score 597; DB 2; Length 590;  
Best Local Similarity 28.9%; Pred. No. 1.2e-36;  
Matches 187; Conservative 111; Mismatches 248; Indels 100; Gaps 18;

Qy 32 GTEARHSLGVLHVSYSVSNRVGPWWNIKSC--QQKWDRQILKDVSLYIESGQIMCILGS 88  
| | : | ::|| : : | : | : | : ||||| | : | | |  
Db 12 GREISYRLETKNLSYRIGGNTPKFSNL--CGLLSEKEEKVILKDVSCDARSAEITAIAGP 69  
  
Qy 89 SGSGKTTLLDAISGRLLRRTGTLEGEVFNVCGLRRDQFQDCFSYVLQSDVFLSSLTVRET 148  
||:|||||: ::||: | : | : | || : :: : | | | |||:|  
Db 70 SGAGKTTLLEILAGKVSH-GKVSQQLVNGRPMGPEYRRVSGFVPQEDALFPFLTQVET 128  
  
Qy 149 LRYTAMLALCRSSADFYNNKKVEAVMTLSLSHVADQMIGSYNFGGISSGERRRVSIQAQL 208  
| |:|:| | | ||: :: || | |||| | | : ||| ||||| | : |  
Db 129 LTYSALLRLKTKRKD-AAAKVKRLIQELGLEHVADSRIGQSGSRSGISGGERRRVSIGVEL 187  
  
Qy 209 LQDPKVMMLDEPTTGLDCMTANQIVLLLAELA-RRDRIVIVTIHQPRSELFQHFDKIAIL 267

```

      : || |::|||:| | :| |:| || :: :: : ::||| | : : | :| :|
Db      188 VHDPNVILIDEPTSGLDASALQVVTLTKDMTIKQGKTIVLTIHQPGFRILEQIDRIVLL 247

Qy      268 TYGELVFCGTPEEMLGFFNNGCGYPCPEHSNPFDFYMD----LTSVDTQSREREIETYKRV 323
      : | :| | : : : | : | | : : | | : || | | | |
Db      248 SNGMVVQNGSVVSLHQIKFSGHQIPRRVNVLEYAIDIAGSLEPIRTQSC-REISCYGH 306

Qy      324 QMLECAF-----KESDIY-HKILENIERARYLKTLPMPFKTKDPPGMFGKLGVLRR 375
      : : : :|| : : :|| :: :| :|
Db      307 KTWKSCYISAGGELHQSDSHSNSVLEEVQ-----ILGQR 340

Qy      376 VTRNLMRNKQAVIMRLVQNLMGLFLIFYLLRVQNTLKGAVQDRVG----LLYQLVGAT 431
      :| : | | | :| | | | | | : | | | :| | : :|
Db      341 SCKNIFRTKQLFTRALQASIAGLILGSIYLVNGNQKKEAKVL-RTGFFAFILTFLLSST 399

Qy      432 PYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYVLHVLFPFSVIATVIFSSVCYWT 491
      : :| | : :| | :|| | :|| :| ::| :| | :
Db      400 -----TEGLPIFLQDRRIIMRETSRRAYRVLSYVLADTLIFIPFLLIISMLFATPVYWL 454

Qy      492 GLYPEVARFGYFS-----AALLAPHLIGEFLTLVLLGIVQNPNIIVNSIVALLSISGLL-- 544
      || | : | || | : : : | || | | | : | : ||| :
Db      455 GLRRELDGFLYFSLVIWIVLLMSNSFVACFSALV-----PNF---IMGTSVISGLMGS 504

Qy      545 --IGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPMCA 602
      : || : :| : : : | : || | | :|| :| |
Db      505 FFLFSGYFIAKDRIPVYWEFMYLSLFKYPFECLMINEYRGDVFL----- 549

Qy      603 ITQGVQFIEKTCPGATSRTANFLILYGFIPALVILGIVIFKVRDY 648
      | | : :| | : || :|| | | |
Db      550 -----KQQDLKESQKWSNLGIMASFIVGYRVLGFFILWYRCY 586

```

# RESULT 7

E96742

probable ABC transporter F17M19.11 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001

C;Accession: E96742

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.; Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.; Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.; Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda, M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken, S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis. <sup>P</sup>

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: E96742  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-609 <STO>  
 A;Cross-references: GB:AE005173; NID:g6978921; PIDN:AAF34313.1; GSPDB:GN00141  
 C;Genetics:  
 A;Gene: F17M19.11  
 A;Map position: 1  
 C;Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 17.7%; Score 595; DB 2; Length 609;  
 Best Local Similarity 29.6%; Pred. No. 1.8e-36;  
 Matches 178; Conservative 121; Mismatches 235; Indels 68; Gaps 17;

Qy	66	DRQILKDVSLYIESGQIMCILGSSSGSKTLLDAISGRLRRTGTLEGEVFNVCCELRRDQ	125
		:      :    :       :    :    :   :::  :::	
Db	27	ERTILSGVTGMISPGEFMAVLGPSGSKSTLLNAVAGRLHGS-NLTGKILINDGKITKQT	85
Qy	126	FQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRS-SADFYNKKVEAVMTELSLSHVADQ	184
		: :     :      :  :      :  :  : :     : :	
Db	86	LKRT-GFVAQDDLLYPHLTVRETLVFVALLRLPRSLTRDVKLRAAESVISELGLTKCENT	144
Qy	185	MIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELAR-RD	243
		:: :        :      :   :  ::   :      :       :	
Db	145	VVGNTFIRGISGGERKRVSIAHELLINPSLLVLDEPTSGLDATAALRLVQTLAAGLAHGKG	204
Qy	244	RIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPFDFYM	303
		:  : :        :      :  :  :    : :  :  :  :  :  :  :  :	
Db	205	KTVVTSIHQPSSRVFQMFDTVLLLESGKCLFVGKGRDAMAYFESVGFSPAFFMNPADFL	264
Qy	304	DLTSVDTQS---RERE-----IETYKRVQMLECAFKESDIYHKILENIERARYL	349
		:  :     :   :   :  :  :  :  :  :  :  :  :	
Db	265	DLANGVCQTDGVTEREKPNVRQTLVTAYDTLLAPQVKTCI----EVSHFPQDN---ARFV	317
Qy	350	KTLPMVPFKTKDPPGMFGKLGVLRLRVTRNLNRNKQAVIMRLVQ---NLIMGLFLIFYL	405
		:  :   : :  : : :  :  :  :  :  :  :  :	
Db	318	KTRVNGGGITTCIATWFSQLCILLHRLLEKE-RRHESFDLLRIFQVVAASILCLMWWHSD	376
Qy	406	LRVQNNTLKGAVQDRVGLLYQLVGATPYTGML---NAVNLFPMLRAVSDQESQDGLYHKW	462
		:  :  :  :  :  :  :  :  :  :  :  :  :	
Db	377	YR-----DVHDRLGLLFFI---SIFWGVLPSEFNAVFTFPQERAI FTRERASGMYT	425
Qy	463	QMLLAYVLHVLPPFSVIATVIFSSVCYWTGLYPEVARFGYFSAALLAPHLIGEFLLVLL	522
		: :    :  :  :  :  :  :  :  :  :  :  :  :  :	
Db	426	SYFMAHVLGSLSMELVLPASFLTFTYWMVYLRPGIVPFLTLSVLLLYVLASQGLGLALG	485
Qy	523	GIVQNPNIIVNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVNEFY	582
		: : :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	
Db	486	AAIMDAKKASTIVTVTMLAFVLTGGYYVNKV---PSGMVWMKYVSTTFYCYRLLVAIQY-	541
Qy	583	GLNFTCGGSNTSMLNHPMC-----AITQGVQFIEKTCPGATSRFTA---NFLILY	629
		:    :  :  :  :  :  :  :  :  :  :  :  :  :	
Db	542	-----GSGEILRMLGCDKSGKQGASAATSAGCRFVEEEVIGDVGWMTSVGVFLFMFF	594
Qy	630	GF	631
		:	
Db	595	GY	596

## FYFFW

C;Species: *Drosophila melanogaster*

C;Accession: S08635; S07263; S10240

Nucleic Acids Res. 18, 1633, 1990

A;Reference number: S08635; MUID:90221897; PMID:2109311

A;Molecule type: mRNA

A;Cross-references: EMBL:X51749; NID:g8825; PIDN:CAA36038.1; PID:g8826

J. Mol. Biol. 180, 437-455, 1984

A:Reference number: S07263; MUID:85134865; PMID:6084717

A;Molecule type: DNA

334. 'ITLHLNSYPAPVPSVLPTTIRRTFTYRCWPLCPDGRSSPVIGSPRYG', 372-687 <OHA1>

A; Experimental source: strain Canton S

submitted to the EMBL Data Library, June 1985

A;Accession: S10240

A;Residues: 1-24, 'LIFEIPYHCRVTAD', 30-687 &lt;OHA2&gt;

A; Experimental source: strain Canton S

A; Gene: white; w

A; Introns: 24/3; 116/1; 334/2; 439/3; 483/3

C;Keywords: ATP; glycoprotein; nucleotide binding; P-loop; transmembrane protein

F:130-137/Region: nucleotide-binding motif A (P-loop)

F:67.93,472,554,651/Binding site: carbohydrate (Asn) (covalent) #status

predicted

Best Local Similarity 27.5%; Pred. No. 6e-36;

Matches 200; Conservative 129; Mismatches 255; Indels 142; Gaps 25;

Db 13 GSKHPSAEHLNNGDSGAASQSCINQGFQAKNYGTLLPPSPPEDSGSGSGQLAENLTYAW 72

Db 73 HNMDI---FGAVNQPGSGWRQLVNRTRGLFCNERHIPAPRKHLLKNVCGVAYPGELLAV 128



Qy 86 LGSSGSGKTTLLDAISGRLRR--TGTEGEVFNVCCELRRDQFQDCFSYVLQSDVFLSSL 143  
 :|||:||||:|:| : : | :|| : : | :|| | |:|: ||  
 Db 129 MGSSGAGKTTLLNALAFRSPQGIQVSPSGMRLNGQPVDAKEMQARCAIVQDDDLFIGSL 188  
 Qy 144 TVRETLRYTAMLALCRSSADFYNK---KVEAVMTELSLSHVADQMIG-SYNFGGSISSGER 199  
 | || | : ||: : | : : :|: | |||| :|| | :| |||  
 Db 189 TAREHLIFQAMVRMPRHLT--YRQVARVDQVIQELSLSKCQHTIIGVPRVKGLSGGER 246  
 Qy 200 RRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQ 259  
 :|:: |:: | || ::: |||:| || :| :| :|:: : ||:| ||| |||:  
 Db 247 KRLAFASEALTDPLLICDEPTSGLDSTAHSVVQVLKKSQKGKTVILTIHQPSSELFQ 306  
 Qy 260 HFDKIAILTYGELVFCGTPEEMLGFFNNGCYPCPEHSNPFDFYMDLTSVDTQSREREIET 319  
 |||| :: | : | || | : ||: | || : || |||: : :| ||||:  
 Db 307 LFDKILLMAEGRVAFLGTPSEAVDFFSYVGAQCPTNYPADFYVQVLAV---VPGREIES 363  
 Qy 320 YKRVQMLECAFKESDIYHKILENIERARYLKTLPMPFKTKDPP-----GMFGKLG 371  
 | : : | | | : :||: | | | : | | :  
 Db 364 RDRIAKICDNFAIS---KVARDEQLLATKNLE---KPLEQPENGYTYKATWFMQFRA 415  
 Qy 372 LLRRVTRNLMRNKQAVIMRLVQNLMIGLFLIFYLLRVQNNTLKGAVQDRVGLLYQLVGAT 431  
 :| | ::: | :||:| : : : || : | | | : | : :  
 Db 416 VLWRSWLSVLKEPLLKVRLIQTMTVAI-LIGLIFLGQQLTQVG-VMNINGAIFLFLTNM 473  
 Qy 432 PYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYVLHVLFPFSVIATVIFSSVCYWT 491  
 : : :|: | :||: || | : || : :|: :| : :  
 Db 474 TFQNVFATINVTSELPVEMREARSRLYRCDTYFLGKTIAELPLFLTVPVFTAIAYPMI 533  
 Qy 492 GLYPEVARFGYFSAALLAPHLIGEFLLTLVLLGIVQNPINVSIVALLS----- 539  
 || | | | | | :|: :| :| :| :|  
 Db 534 GLRAGVLHF-----FNCLALVTLV--ANVSTSFGLYISCASSSTSMALSV 576  
 Qy 540 ----ISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVNEFYGL---NFTCGGSN 592  
 | | :| | | :|: || | | : :| | | :|: : : :| ||  
 Db 577 GPPVVIIPFLFEGGFFLNSGSPVYLKWLSTWFRYANEGLLINQWADVEPGEISCTSSN 636  
 Qy 593 TSMLNHPMCAITQGVQFIEKTCPGA-----TSRFTANFLILYGFIPALVILGIVIFKVR 646  
 | ||| : | | :| | | :| | | | :|  
 Db 637 T-----TCPSSGKVILETLNFSADLPL-DYV-GLAIL-IVSFRVL 674  
 Qy 647 DYLISS 652  
 || |  
 Db 675 AYLALR 680

RESULT 9

T47650

ABC transporter-like protein - Arabidopsis thaliana

N;Alternate names: protein T15C9.110

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 19-May-2000

C;Accession: T47650

R;Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.

submitted to the Protein Sequence Database, April 2000

A;Reference number: Z24470

A;Accession: T47650

A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-708 <MEW>  
 A;Cross-references: EMBL:AL132970  
 A;Experimental source: cultivar Columbia; BAC clone T15C9  
 C;Genetics:  
 A;Map position: 3  
 A;Note: T15C9.110  
 C;Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein  
 Fl2L6.1; ATP-binding cassette homology

Query Match 17.4%; Score 587; DB 2; Length 708;  
 Best Local Similarity 27.8%; Pred. No. 8.9e-36;  
 Matches 159; Conservative 129; Mismatches 229; Indels 54; Gaps 15;

```

Qy      67 RQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVCGLRRDQF 126
      : :| |:: |:: :|| ||::||::||::||: |::| | :|| :: : :
Db      89 KTLDDITGEARDGEILAVLGSGAGKSTLIDALAGRAE--DSLKGTVTLNKEKVLQSRL 147

Qy     127 QDCFS-YVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMT---ELSLSHVA 182
      | ||:| |:: ||::|| : : | || :||:| | | :| | : |
Db     148 LKVISAYVMQDDLLFPMLTVKETLMFASEFRLPRSLPK--SKKMERVETLIDQLGLRNAA 205

Qy     183 DQMIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARR 242
      | :|| |::| ||||| |:: || : : ||||:| | | :| :| :|:
Db     206 DTVIGDEGHRGVSGGERRRVSIGIDIHDPILLFLDEPTSGLDSTNAFMVVQVLKRIAQS 265

Qy     243 DRIVIVTIHQPRSELFQHFDKIAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPFDFY 302
      :||::||| : : |:: ||::|: || |:| : ||:| | | | | :|
Db     266 GSVVIMSIHQPSARIIGLLDRLIILSHGKSVFNGSPVSLPSFFSSFGRIPEKENITEFA 325

Qy     303 MDLTS-----VDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYL- 349
      :|: |:: |:: : : :| : : ||: | : : | : :
Db     326 LDVIRELEGSSEGTRDLVEFNKQWQONQTARATTQSRVSLKEA-----IAASVSRGKLVS 380

Qy     350 -----KTLPMVPFKT-KDPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIF 403
      : | : :|| : :| :| :| :| : : ||: : : || |
Db     381 GSSGANPISMETVSSYANPP--LAETFILAKRYIKNWIRTPELIGMRIGTVMVTGLLLAT 438

Qy     404 YLLRVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQ 463
      | : :|| :|| |::| : : :| : : :| | : :|: |
Db     439 VYWRL-DNTPRGA-QERMG-FFAFGMSTMFYCCADNIPVFIQERYIFLRETHNAYRTSS 495

Qy     464 MLLAYVLHVLFPFSVIATVIFSSVCYWTGLYPEVARFGYFSAALLAPHLIGEFLLVLLG 523
      :::: | || : : : ||: ||::| : | | : : | | : : |
Db     496 YVISHALVSLPQLLALSIAFAATTFTWTVGLSGGLESFFYYCLIIYAAFWSGSSIVTFISG 555

Qy     524 IVQNPNIIVNS-IVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFY 582
      : : ||:: | :| : :| | : | | :| : | | | :::|||
Db     556 LI--PNVMSYMTIAYLSYCLLLGGFYINRDRIPLYWIWFHYISLLKYPYEAVLINEF- 612

Qy     583 GLNFTCGGSNTSMLNHPMCAITQGVQFIEKT 613
      : | :||| : |
Db     613 -----DDPSRCFVKGVQVFDGT 629
  
```

RESULT 10

G02068  
white homolog - human  
C;Species: Homo sapiens (man)  
C;Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 02-Feb-2001  
C;Accession: G02068  
R;Croom, J.M.; Tiller, G.; Fletcher, J.A.; Lux, M.; Raab, E.; Goldenson, D.;  
Arciniegas, S.; Son, D.; Wu, R.  
submitted to the EMBL Data Library, August 1995  
A;Reference number: H00769  
A;Accession: G02068  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-638 <CRO>  
A;Cross-references: EMBL:U34919; NID:g1314276; PIDN:AAC51098.1; PID:g1314277  
C;Genetics:  
A;Gene: white  
C;Superfamily: fruit fly white protein; ATP-binding cassette homology  
C;Keywords: ATP; nucleotide binding; P-loop  
F;61-253/Domain: ATP-binding cassette homology <ABC>  
F;78-85/Region: nucleotide-binding motif A (P-loop)

Query Match 17.2%; Score 580; DB 2; Length 638;  
Best Local Similarity 28.1%; Pred. No. 2.6e-35;  
Matches 164; Conservative 125; Mismatches 241; Indels 54; Gaps 15;

Qy	23	SSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKSCQOKWDRQILKDVSLYIESGQI	82
		: :     :           :   :     :	
Db	27	SSLPRRAAVNIEFR-----DLSYSVPE--GPWW-----RKKGYKTLKLGISGKFNSGEL	73
Qy	83	MCILGSSSGSGKTTLLDAISGRLLRRTGTLEGEVFNVCCELRRDQ--FQDCFSYVLQSDVFL	140
		:   :       :   :   :   :       :     :     :   :     :	
Db	74	VAIMGPSGAGKSTLMNLAG-YRETG-MKGAVLING--LPRDLRCFRKVCYIMQDDMLL	129
Qy	141	SSLTVRETLYRTAMALCRSSADFYNNKKEAVMTLSLSHVADQMIGSYNFGGISSGERR	200
		:   : :     : : :   : :       :     :     :	
Db	130	PHLTVQEAMMVSAHLKL-QEKDEGRREVMKEILTALGLLSCANTRTGS-----LSGGQRK	183
Qy	201	RVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQH	260
		: :   :   :           :   :   :   :   :       :   :	
Db	184	RLAIALELVNPPVPMFFDEPTSGLDASCFQVVSIMKGLAQGGRSIICTIHPQSAKLFEL	243
Qy	261	FDKIAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETY	320
		: : :   :   :   : : :     :         : :   : : :	
Db	244	FDQLYVLSQGCVCYRGKVCNLVPYLRDLGLNCPTYHNPADFVMEVASGEYGDQNSRLVRA	303
Qy	321	KRVQMLECAFKES-----DIYHKILENIERARYLKTLPMPVFKTKDPPGMFG----	367
		:   : : :   : : :	
Db	304	VREGMCDSDHKRDLGGDAEVNPFLLWHRPSEEVKQTKRLKGL-----RKDSSSMEGCHSF	357
Qy	368	-----KLGVLRLRVTRNLMRNKQAVIMRLVQNLMGLFLIFYLLRVQNNLKGAVQDRV	421
		: :   :   : : : : : : : : :   :	
Db	358	SASCLTQFCILFKRTFLSIMRDSVLTHLRITSHIGIGLLIGLLYLGIGNEAKK--VLSNS	415
Qy	422	GLLYQLVGATPYTGMLNAVNLFPMRLRAVSDQESQDGLYHKWQMLLAYVLHVLPPFSVIATV	481
		: : : :     :   :   :   : :   :	
Db	416	GFLFFSMLFLMFAALMPTVLTFFLEMGVFLREHLNYWYSLKAYYLAKTMADVFPQIMFPV	475

Qy 482 IFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGEFLLTLVLLGIVQNPNIIVNSIVALLSIS 541  
 : |: || : | |: | |: : | | |: | : | ::  
 Db 476 AYCSIVYWMTSQPSDAVAFVLFAALGTMTSLVAQSLGL-LIGAASTSLQVATFVGPVTAI 534  
 Qy 542 GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLN 585  
 : |: ||| : :| |: : | :: :| | :::: |||:  
 Db 535 PVLLFSGFFVSEDTIPTYLQWMSYISYVRYGFEGVILS-IYGLD 577

RESULT 11

T31958

hypothetical protein F02E11.1 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 31-Jan-2000

C;Accession: T31958

R;Favello, A.; Scheet, P.

submitted to the EMBL Data Library, July 1997

A;Description: The sequence of *C. elegans* cosmid F02E11.

A;Reference number: Z21104

A;Accession: T31958

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-658 <FAV>

A;Cross-references: EMBL:AF016661; PIDN:AAB66050.1; GSPDB:GN00020; CESP:F02E11.1

A;Experimental source: strain Bristol N2; clone F02E11

C;Genetics:

A;Gene: CESP:F02E11.1

A;Map position: 2

A;Introns: 115/3; 158/3; 214/3; 330/3; 368/2; 448/3; 525/1

C;Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 17.2%; Score 579.5; DB 2; Length 658;  
 Best Local Similarity 27.8%; Pred. No. 2.9e-35;  
 Matches 169; Conservative 117; Mismatches 254; Indels 69; Gaps 13;

Qy 73 VSLYIESGQIMCILGSSSGSKTLLDAISGRLLRRTGT-LEGEVVFVNGCELRRDQFQDCFS 131  
 || | |::: ::| ||:|||||:: | | | |::| ||| :: : : : :  
 Db 79 VSGVAEPGEVLALMGSGAGKTTLMN-ILAHLDTNNGVEYLGDTVNGKKITKQKMRQMCA 137  
 Qy 132 YVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKEAVMTELSLSHVADQMIGSYN- 190  
 || | |::| :|||| | ||| : : :: :|| |::| : :|| |  
 Db 138 YVQQVDLFCGTLTVREQLTYTAHMRMKNATVQQKMERVENVLRDMNLTDQNTLIGIPNR 197  
 Qy 191 FGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTI 250  
 ||| ||::|:: | ::| |||:: |||:|| |::| |::| : : :|| :  
 Db 198 MKGISIGKKRLAFACEILTDPKILFCDEPTSGLDAFMASEVVRALLDLANKGKTIIVVL 257  
 Qy 251 HQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNCG--YPCPEHSNPFDFYMDLTSV 308  
 ||| |::| |:: : |::| : : ||: | : || |||| | |:  
 Db 258 HQPSSTVFRMFHKVCFMATGKTVYHGAVDRLCPPFDKLGPDFRVPESYNPADFVMSSEISI 317  
 Qy 309 DTQSREREIETKYRVQMLECAFKESDIYHKILENIERARYLKTLPMPVFKTKDPPGMFGK 368  
 | | | | |:: | :: ||| ::| : ||: ||  
 Db 318 ---SPETEQEDVTRIEYLIHEYQNSDIGTQMLK-----KTRTAVDEFEGG 358  
 Qy 369 LG-----VLLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQ 409  
 | :||:| | | : : :| | : : ||:

Db	359	YGDEDDGESRYNSTFGTQFEILLKRSRLRTTFRDPLLLRVRFQAQILATAILVGIVNWRVE	418
Qy	410	NNTLKG-AVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAY	468
		: :  ::  : :  :   :  : :	
Db	419	---LKGPTIQNLEGVMYNCARDMTFLFYFPSVNVITSELPVFLREHKSNIYSVEAYFLAK	475
Qy	469	VLHVLPPFSVIATVIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFLLTVLLGIVQNP	528
		: : ::             : : :     :	
Db	476	SLAELPQYTILPMIYGTIIYWMAGLVASVTSFLVFVFCITLTWVAVSIAYVGACIFGDE	535
Qy	529	NIVNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCAILVNEFYGLNFTC	588
		:  : : : :   :        : : : : : :   :      :: ::	
Db	536	GLVVTFMPMFVLPMLVFG-GFYVNANSIPVYYQYVSFVSWFKHGFEALEANQWKEIDKIS	594
Qy	589	GGNNTSMLNHPMCAITQGVQFIEKTCP-----GATSRFTANFLILYGFIPALVI	637
		: :            :        :	
Db	595	GCD----LINPLNATTGY-----CPASDGPGLTRRGIDTPLYANVLILEMSFFVYRI	644
Qy	638	LGIVIFKVR	646
		: :   :	
Db	645	IGLVALKIR	653

RESULT 12

T47648

ABC transporter-like protein - *Arabidopsis thaliana*

N;Alternate names: protein T15C9.80

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 20-Apr-2000 #sequence revision 20-Apr-2000 #text change 19-May-2000

C;Accession: T47648

R;Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.

submitted to the Protein Sequence Database, April 2000

A;Reference number: Z24470

A;Accession: T47648

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-720 <MEW>

A;Cross-references: EMBL:AL132970

A;Experimental source: cultivar Columbia; BAC clone T15C9

C; Genetics:

```
A;Map position: 3
```

A;Note: T15C9.80

C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein

F12L6.1; ATP-binding cassette homology

Query Match 17.1%; Score 577; DB 2; Length 720;  
Best Local Similarity 27.8%; Pred. No. 5.1e-35;  
Matches 173; Conservative 120; Mismatches 237; Indels 92; Gaps 18;

```

Qy      43 HVSYSVSNR-----VGPWWNIKSCQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTT 95
      ::||:| | : || : | :| ::| | :| :||| |:|
Db      61 NLTYNVSVRRKLDHDLVPWRRSTSFSKTK--TLLDNISGETRDGEILAVLGASGSGKST 117

Qy      96 LLDAISGRLRRTGTLEGEVFNVCCELRRDQFQDCFSYVLQSDVFLSSSLTVRETLRYTAML 155
      |:|::| : : |:|:| | :|| |: : :||:| |: ||| ||| : |
Db     118 LIDALANRIAK-GSLKGTVTLNGEALQSRMLKVISAYVMQDDLFPMLTVEETLMFAAEF 176

```

```

Qy      156 ALCRSSADFYNK-KVEAVMTELSLSHVADQMIGSYNFGGISSGERRRVSIAAQLLQDPKV 214
      | ||      | :|:|: :| : : | :||      ||| ||||| : : || |
Db      177 RLPRSLPKSKKKLRVQALIDQLGIRNAAKTIIGDEGHRGISGGERRRVSIGIDIIHDPV 236

Qy      215 MMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQHFDKIAILTYGELVF 274
      : |||||:| | :| :| :| :| :|:|:| | : | :| :| |
Db      237 LFLDEPTSGLDSTSAFMVVKVLKRIAESGSIIIMSIHQPSHRVLSLLDRLIFLSRGHTVF 296

Qy      275 CGTPEEMLGFFNCGYPCPEHSNPFDFYMDLTSVDTQSRERE-----IETYKRVQM 325
      |:| : || | | | :| :| | | :| :| |
Db      297 SGSPASLPSFFAGFGNPIPENENQTEFALDI-----RELEGSAGGTRGLVEFNKKWQE 350

Qy      326 LECAFKESD-----IYHKILENIERARYLK-----TLPMPV 356
      : : | :| : : | :| :| : : | :| :
Db      351 MK---KQSNPQTLTPPASPNPNLTKEAISASISRGKLVSGGGGGSSVINHGGGT LAVPA 407

Qy      357 FKTKDPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLMGLFL--IFYLLRVQNNTLK 414
      | : : : | | | : : ||| : : | | :| :| :| :|
Db      408 FANP----FWIEIKTLTRRSILNSRRQPELLGMRLATVIVTGFILATVFWRL---DNSPK 460

Qy      415 GAVQDRVGLLYQLVGATPYTGMLNAVNLFPMRAVSDQESQDGLYHKWQMLLAYVLHVLP 474
      | ||:|:| : :| : :| : :| | : :| : | : :| : : |
Db      461 G-VQERLG-FFAFAMSTMFYTCADALPVFLQERYIFMRETAYNAYRRSSYVLSHAIVTFP 518

Qy      475 FSVIATVIFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPNIIV--- 531
      : : :| : :| :| : | : : :| | | | :| | : :
Db      519 SLIFLSLAFAVTTFWAVGLEGLMGFLFYCLIILASFWSGSSFVTFLSGVV--PHVMLGY 576

Qy      532 NSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGS 591
      :||:|: | : ||| | :| | : || | : : |||
Db      577 TIVVAILAY--FLLFSGFFINRDRIPOYWIWFHYLSLVKYPYEAVLQNEF----- 624

Qy      592 NTSMLNHPMCAITQGVQFIEKT 613
      : | :|| : :
Db      625 -----SDPTECFVRGVQLFDNS 641

```

# RESULT 13

T34391

hypothetical protein T26A5.1 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 04-Mar-2000

C;Accession: T34391

R;Du, Z.

submitted to the EMBL Data Library, April 1994

A;Description: The sequence of *C. elegans* cosmid T26A5.

A;Reference number: Z21516

A;Accession: T34391

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-608 <DUZ>

A;Cross-references: EMBL:U00043; PIDN:AAC77504.1; GSPDB:GN00021; CESP:T26A5.1

A;Experimental source: strain Bristol N2; clone T26A5

C;Genetics:

A;Gene: CESP:T26A5.1

A;Map position: 3

A;Introns: 23/1; 96/3; 243/1; 342/2; 374/3; 403/1; 428/2; 464/3; 494/3; 534/2

C;Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 17.0%; Score 573.5; DB 2; Length 608;  
Best Local Similarity 26.2%; Pred. No. 7.4e-35;  
Matches 161; Conservative 118; Mismatches 249; Indels 87; Gaps 14;

```
Qy      64 KWDRQILKDVSLYIESGQIMCILGSSSGSKTLLDAISGRLRRTGTLEGEVFNVCCELRR 123
      | | :||:| | :||:: :||:| | | | | : : | : : || : || | | :
Db      40 KEKRLLLKNVSGYAKSGELLALMGASGAGKTLLNMLMCRNLKGLSTEGTITVNGNEMAH 99

Qy     124 DQFQDCFSYVLQSDVFLSSLTVRETLYTAMLALCRSSADFYNNKVEAVMTELSLSHVAD 183
      : : | :||: | | | | : | : | : | | : | | | | |
Db     100 -KISSISGFAQQEELFVGTTLTVKEYLMIQAKLRI-NGSKKLREDRVTDVLHQLKLWKCRD 157

Qy     184 QMIGSY-NFGGSISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARR 242
      || | | | | | | | | : | : | : | : | | | | | | | | : : : | : :
Db     158 SKIGVIGEEKKGISGGEARRLTFACEMLSNPSSLFADEPTTGLDSFMAESVIQILKGIKT 217

Qy     243 DRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNGGYPCPEHSNPFDFY 302
      | : | | | | | : | : | | : | | | | : | | | | : | : | | : :
Db     218 GRTIICTIHPSSQLYQMFHRVIYLANGSTAFQGTPESESISFFEKCGHRVPDEYNPSEWI 277

Qy     303 MDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPVFPKTKDP 362
      : : | : | : | : | : : : | : : : | : : | : | |
Db     278 IYKLAVQP---GQEKQSNDRIQKIVEQYEDSDHQKRVMEQLS-----DVSEKIP 323

Qy     363 P-----GMFGKLGVLRLRRVTRNLNRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNTLKGA 416
      | : : | : | : : | : | : : | : : | : | : | : | : | :
Db     324 PPEMHRANVFTQIFALSTRCGIDVWRAPQLTLAKVIQKILFGLFIGLLYLRTPYDA--RG 381

Qy     417 VQDRVGLLYQLVG----ATPYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYVLHV 472
      : : | | : | | : | | | | | : : : | | | : | | : :
Db     382 IHNINGALFFLAGIYIYSTAYAIMFFLNNEFPLVA----REYHDGLYNLWTFYFARCISL 437

Qy     473 LPFSVIATVIFSSVCYWTGLGYPEV-----ARFGYFSAALLAPHLIGE 515
      : | : | : | : | | | | : | | : :
Db     438 IPLFSTDGLILLFIVYWLIGLNTSVMQVIVASIITVLASQAASAFGIAMSCI----- 489

Qy     516 FLTLVLLGIVQNPINVSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEI 575
      | | : : : | : | : | | | | | : : | : : : : |
Db     490 FPTAQMTAVMASPPLV-----LFRLEFGGLYG-----NTNTFPAAIRWLQWISMYRFAFEG 539

Qy     576 LVVNEFYGLNFTCGGSNTSMLNHPMCAITQGVQFIEKTCPGATSRFTANFLILYGFIPAL 635
      | | | : : : : | : : | | | |
Db     540 LVVNQWSEIDDFHSNAKNNWTNSTNDVLDYFAFSESAIP-----L 580

Qy     636 VILGIVIFKVRDYLI 650
      | : | : : : | | |
Db     581 DIIGLILISLAFYLI 595
```

RESULT 14

D96553

hypothetical protein F5D21.6 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001

C;Accession: D96553

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;  
Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.;  
Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.;  
Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.;  
Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.;  
Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.

A; Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzialli, A.; Militscher, J.; Miranda, M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken, S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: D96553

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-687 &lt;STO&gt;

A;Cross-references: GB:AE005173; NID:g10092349; PIDN:AAG12758.1; GSPDB:GN00141

C; Genetics:

A; Gene: F5D21.6

```
A;Map position: 1
```

C;Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein

F12L6.1; ATP-binding cassette homology

Query Match 17.0%; Score 571.5; DB 2; Length 687;

Best Local Similarity 28.5%; Pred. No. 1.2e-34;

Matches 160; Conservative 120; Mismatches 216; Indels 65; Gaps 16;

QY           67 RQILKDVSLYIESGQIMCILGSSSGSKTTLLDAISGLRRTGTLEGEVFNVCCELRRDQF 126  
             |::| : : | :||| :||| ::||| : :| : :| :|

Db 42 RRLDGLNGHAEPGRIMAIMGPSGSGKSTLLDSL AGR LARNVIMTGNLLNGKKARLD-- 99

QY 127 QDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADF----YNKKVEAVMTELSLSHVA 182

Db 100 YGLVAYVTOEDILMGTTLVRETITYSAHL---RLSSDLTKEEVNDIVEGTIIELGLQDCA 156

Ov 183 DQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELAR- 241

[illegible]

242 RDDIVIVTTHORRSELEFOHDKIAILLTYGELVECGTPEEMLGFEFNNCGYPCBEHSNPEDF 301

QY 242 RDRIVIVV11HQPRSELFQHFDRIATLETFGEELVFCGTFEELMGGTFNNCGTFCELEASNTYDF 501  
| | : : |||| | | : | | | : : | : | | : | | : : | | : | | : | | : | |

Db 217 GGRTVVSSIHQPSSEVFALEDDLFLSSSGEIVYIFGESKFAVEFFAEAGFFCFARRANFSDH 270

Qy 302 YMDLTSVDLTQSRERELETYKRVQMLECA-----FRESDIYHRILENTERARILEKILFMV 355  
 :: : | : :: : | :: | : | :: | | : | | :

Db 277 FLRCINSDFDTVTATLKGSQRIRETPTSDDLMLNLATSEIKARLVENYRRSVYAKS---A 333

Qy 356 PFKTKDPPGMFGKLGVLRLRVTRNLNRNKQAVIMRLVQNLIMGLF-----LIFYLLRVQ 409

Db 334 KSRIRELASIEGHHGMEVR-----KGSEATWFKQLRTLTKRSFVNMCRDIGYYWSRI- 385



Qy 410 NNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNL-----FPML---RAVSD 451  
 : | || :: || : || : | : || |  
 Db 386 --VIYIVVSFCVGTIFYDVGHs-YTSILARVSCGGFITGFMTFMSIGGFPSFIEEMKVfY 442

Qy 452 QESQDGLYHKWQMLLAYVLHVLPPFSVIATVIFSSVCYWTGLGLYPEVARFGYFSAALLAPH 511  
 :| | | : :: : || | :| | : | : : :| :  
 Db 443 KERLSGYYGVSVYIISNYVSSFPFLVAIALITGSITYNMVKFRPGVSHWAFFCLNIFFSV 502

Qy 512 LIGEFLLVLLGIVQNPNIvNSIVALLSISG-LLIGSGFIRNIQEMPIPLKI-----LGY 565  
 : | | :| : :| || : :: | | :: || | : ::| | : : :  
 Db 503 SVIESIMMVVASLV--PNFLMGLITGAGIIGIIMMTSGFFRLLPDLP---KVEWRYPIsF 557

Qy 566 FTFQKYCCEILVVNEFYGLNF 586  
 :: : : | :| || |  
 Db 558 MSYGSWAIQGAYKNDFLGLEF 578

# RESULT 15

B88474

protein C05D10.3 [imported] - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 15-Jun-2001

C;Accession: B88474

R;anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology.

A;Reference number: A75000; MUID:99069613; PMID:9851916

A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and

www\_sanger.ac.uk/Projects/C\_elegans/ for a list of authors

A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and Science 285, 1493, 1999

A;Accession: B88474

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-559 <STO>

A;Cross-references: GB:chr\_III; PIDN:AAA20989.1; PID:g532111; GSPDB:GN00021;

CESP:C05D10.3

A;Note: similar to D. melanogaster white protein

C;Genetics:

A;Gene: C05D10.3

A;Map position: 3

C;Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 16.9%; Score 570; DB 2; Length 559;

Best Local Similarity 25.6%; Pred. No. 1.2e-34;

Matches 150; Conservative 125; Mismatches 241; Indels 70; Gaps 13;

Qy 67 RQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVCCELRRDQF 126  
 :||| :|| |||::: |||||:|||||:: : : | :|| : :|| : :  
 Db 7 KEILHNVSgMAESGKLLAILGSSGAGKTTLMNVLTsrNLtnLDVQGSILIDGRRANKWKI 66

Qy 127 QDCFSYVLQSDVFLSSLTVRETlRYTAMlALCRSSADfYNK-----KVEAVMTelSLSHV 181  
 :: :|| | | :| : :| || | : : | | :| : :|| | :| : :|  
 Db 67 REMSAFVQQHDMFVGTMtAREHLQFMARL---RMGDQYYSdHERQLRVEQVLTQMGLKKC 123

Qy 182 ADQMIGSYN-FGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240  
 || :|| | :| ||::|| |::| ||:: |||:| | :| | ||  
 Db 124 ADTVIGIPNQLKGLSCGEKKRLSFASEILTCPKILFCDEPTSGLD AFMAGHV VQALRSLA 183

Qy 241 RRDRIVIVTIHQPRSELFQHFDKIAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPF 300  
 ||:|||| | :: | : : : | :: | : : | ||||| : || |  
 Db 184 DNGMTVIITIHQPSSHVYSLFNNVCLMACGRVIYLGPGDQAVPLFEKCGYPCPAYNPAD 243

Qy 301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLP MV----- 355  
 : :| | ::| :: | :| :| | | : |  
 Db 244 HLIRTLAVIDSDRATSMKTISKIRQ---GFLSTDLGQSVLA-IGNANKLRAASFVTGSDT 299

Qy 356 -----PFKTKDPPGMF-GKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLF--LIFYLLR 407  
 | :| | : | | ::| : :||:| | :||  
 Db 300 SEKTKTFFNQDYNASFWTQFLALFWRSWLTVIRDPNLLSVRLLQILITAFITGIVFF--- 356

Qy 408 VQNNTLKGA VQDRVGLLYQLVGATPYTGMLNAVNLFPM LRAVSDQ-----ESQDGLYHK 461  
 | : |:: : :| : || : : : | : :|  
 Db 357 -QTPVTPATIISINGIMFN-----HIRNMNFM LQFPNVPVITAE LPIVLRENANGVYRT 409

Qy 462 WQMLLAYVLHVL PFSVIATVIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEF LTLVL 521  
 || : || :| :::: || ||| : : | : : :  
 Db 410 SAYFLAKNIAELPQYIILPILYNTIVYWM SGLYPNFWNYCFASLV TILITNVAISISYAV 469

Qy 522 LGIVQNPNI VNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEF 581  
 | | :: :| : : : | | :| | | :|| | | :||:  
 Db 470 ATIFANTDVAMTILPIFVVPIMAFG-GFFITFDAIPSYFKWLSSLSYFKYGYEALAIN EW 528

Qy 582 YGLNFTCGGSNTSMLNHPMCAITQGVQFIEKTCPGATSRFTANFLI 627  
 : | | : || ||  
 Db 529 DSI-----KVIPESISRLRIKFLI 547

Search completed: February 27, 2004, 07:18:52  
 Job time : 16.5272 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2004, 07:17:39 ; Search time 29.3006 Seconds  
(without alignments)  
4698.604 Million cell updates/sec

Title: US-09-989-981A-2  
Perfect score: 3369  
Sequence: 1 MGELPFLSPEGARGPHINRG.....PALVILGIVIFKVRDY LISR 652

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

8  
Result Query  
No. Score Match Length DB ID Description

1	3369	100.0	652	9	US-09-837-992-1	Sequence 1, Appli
2	3369	100.0	652	10	US-09-989-981A-2	Sequence 2, Appli
3	2744.5	81.5	651	9	US-09-837-992-3	Sequence 3, Appli
4	2744.5	81.5	651	10	US-09-989-981A-6	Sequence 6, Appli
5	2744.5	81.5	651	14	US-10-090-455-6	Sequence 6, Appli
6	1177	34.9	256	15	US-10-104-047-2795	Sequence 2795, Ap
7	701.5	20.8	672	10	US-09-989-981A-4	Sequence 4, Appli
8	693.5	20.6	655	10	US-09-961-086-1	Sequence 1, Appli
9	693.5	20.6	655	15	US-10-405-806-13	Sequence 13, Appl
10	691.5	20.5	655	9	US-09-981-353-35	Sequence 35, Appl
11	691.5	20.5	655	14	US-10-120-687-61	Sequence 61, Appl
12	691.5	20.5	655	15	US-10-405-806-2	Sequence 2, Appli
13	689.5	20.5	655	9	US-09-866-866A-10	Sequence 10, Appl
14	689.5	20.5	655	14	US-10-090-455-5	Sequence 5, Appli
15	688.5	20.4	673	10	US-09-989-981A-8	Sequence 8, Appli
16	688.5	20.4	673	14	US-10-090-455-7	Sequence 7, Appli
17	683.5	20.3	655	9	US-09-866-866A-27	Sequence 27, Appl
18	677	20.1	657	9	US-09-866-866A-14	Sequence 14, Appl
19	628	18.6	1095	15	US-10-369-493-2025	Sequence 2025, Ap
20	602.5	17.9	1049	15	US-10-369-493-1520	Sequence 1520, Ap
21	592.5	17.6	674	14	US-10-090-455-4	Sequence 4, Appli
22	592.5	17.6	674	16	US-10-429-160-10	Sequence 10, Appl
23	586	17.4	663	13	US-10-108-605-245	Sequence 245, App
24	580	17.2	638	13	US-10-072-621-10	Sequence 10, Appl
25	579.5	17.2	658	15	US-10-369-493-5347	Sequence 5347, Ap
26	576.5	17.1	646	13	US-10-072-621-9	Sequence 9, Appli
27	576.5	17.1	646	14	US-10-090-455-2	Sequence 2, Appli
28	574.5	17.1	599	15	US-10-210-130-14	Sequence 14, Appl
29	573.5	17.0	608	15	US-10-369-493-5748	Sequence 5748, Ap
30	570	16.9	559	15	US-10-369-493-5740	Sequence 5740, Ap
31	569.5	16.9	627	14	US-10-090-455-8	Sequence 8, Appli
32	569	16.9	604	9	US-09-745-763-197	Sequence 197, App
33	562.5	16.7	646	13	US-10-154-452-4	Sequence 4, Appli
34	558.5	16.6	646	14	US-10-079-087-2	Sequence 2, Appli
35	555.5	16.5	646	14	US-10-090-455-13	Sequence 13, Appl
36	554.5	16.5	646	13	US-10-154-452-8	Sequence 8, Appli
37	554	16.4	610	15	US-10-369-493-5687	Sequence 5687, Ap
38	526.5	15.6	676	15	US-10-369-493-3799	Sequence 3799, Ap
39	523.5	15.5	639	15	US-10-369-493-6184	Sequence 6184, Ap
40	520	15.4	695	15	US-10-369-493-6199	Sequence 6199, Ap
41	487.5	14.5	545	14	US-10-083-357-1335	Sequence 1335, Ap
42	480	14.2	551	15	US-10-369-493-3562	Sequence 3562, Ap
43	476.5	14.1	560	15	US-10-369-493-12899	Sequence 12899, A
44	454.5	13.5	615	10	US-09-949-029-24	Sequence 24, Appl
45	423	12.6	1549	15	US-10-369-493-3919	Sequence 3919, Ap

#### ALIGNMENTS

RESULT 1  
 US-09-837-992-1  
 ; Sequence 1, Application US/09837992  
 ; Patent No. US20020081687A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tian, Hui

```

; APPLICANT:  Schultz, Joshua
; APPLICANT:  Shan, Bei
; APPLICANT:  Tularik Inc.
; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG):  Compositions
; TITLE OF INVENTION:  and Methods of Use
; FILE REFERENCE: 018781-006020US
; CURRENT APPLICATION NUMBER: US/09/837,992
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/198,465
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 60/204,234
; PRIOR FILING DATE: 2000-05-15
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 652
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: mouse sitosterolemia susceptibility gene (SSG)
; OTHER INFORMATION: amino acid sequence
US-09-837-992-1

```

```

Query Match          100.0%;  Score 3369;  DB 9;  Length 652;
Best Local Similarity 100.0%;  Pred. No. 0;
Matches 652;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

Qy      1 MGELPFLSPEGARGPHINRGSLSSEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS 60
        |||
Db      1 MGELPFLSPEGARGPHINRGSLSSEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS 60

Qy     61 CQQKWDRQILKDVSLEYIESGQIMCILGSSSGSKTTLLDAISGRLRRTGTLEGEVFNVC 120
        |||
Db     61 CQQKWDRQILKDVSLEYIESGQIMCILGSSSGSKTTLLDAISGRLRRTGTLEGEVFNVC 120

Qy    121 LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKEAVMTLSLSH 180
        |||
Db    121 LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKEAVMTLSLSH 180

Qy    181 VADQMIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240
        |||
Db    181 VADQMIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240

Qy    241 RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPF 300
        |||
Db    241 RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPF 300

Qy    301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPFKTK 360
        |||
Db    301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPFKTK 360

Qy    361 DPPGMFGKLGVLRLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNLTGKAVQDR 420
        |||
Db    361 DPPGMFGKLGVLRLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNLTGKAVQDR 420

Qy    421 VGLLYQLVGATPYTGMLNAVNLFPMRLRAVSDQESQDGLYHKWQMLLAYVLHVLFPFSVIAT 480
        |||

```

```

Db      421 VGLLYQLVGATPYTGMLNAVNLFPMRLRAVSDQESQDGLYHKWQMLLAYVLHVLFPFSVIAT 480
Qy      481 VIFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 VIFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI 540
Qy      541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCCGSNTSMLNHPM 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCCGSNTSMLNHPM 600
Qy      601 CAITQGVQFIEKTCPGATSRFTANFLILYGFIPALVILGIVIFKVRDYLISR 652
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 CAITQGVQFIEKTCPGATSRFTANFLILYGFIPALVILGIVIFKVRDYLISR 652

```

RESULT 2

US-09-989-981A-2

```

; Sequence 2, Application US/09989981A
; Publication No. US20030049730A1
; GENERAL INFORMATION:
; APPLICANT: Hobbs, Helen H.
; APPLICANT: Shan, Bei
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 652
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: mouse ABCG5 (mABCG5)
US-09-989-981A-2

```

```

Query Match          100.0%;  Score 3369;  DB 10;  Length 652;
Best Local Similarity 100.0%;  Pred. No. 0;
Matches 652;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

Qy      1 MGELPFLSPEGARGPHINRGSLSLEQGSVTGTTEARHSLGVLHVSYSVSNRVGPWWNIKS 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MGELPFLSPEGARGPHINRGSLSLEQGSVTGTTEARHSLGVLHVSYSVSNRVGPWWNIKS 60
Qy      61 CQQKWDRQILKDVSLYIESGQIMCILGSSSGSKTTLLDAISGRLRRTGTLEGEVFNVCCE 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CQQKWDRQILKDVSLYIESGQIMCILGSSSGSKTTLLDAISGRLRRTGTLEGEVFNVCCE 120
Qy      121 LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTELSLSH 180

```

```

Db      121 LRRDQFQDCFSYVLQSDVFLSSLTVRETLYRTAMLALCRSSADFYNNKKVEAVMTELSLSH 180
Qy      181 VADQMIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240
Db      181 VADQMIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240
Qy      241 RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPF 300
Db      241 RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPF 300
Qy      301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPVFPKTK 360
Db      301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPVFPKTK 360
Qy      361 DPPGMFGKLGVLRLRRVTRNLNRNKAQAVIMRLVQNLIMGLFLIFYLLRVQNNLTGKAVQDR 420
Db      361 DPPGMFGKLGVLRLRRVTRNLNRNKAQAVIMRLVQNLIMGLFLIFYLLRVQNNLTGKAVQDR 420
Qy      421 VGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYVLHVLFPFSVIAT 480
Db      421 VGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYVLHVLFPFSVIAT 480
Qy      481 VIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI 540
Db      481 VIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI 540
Qy      541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPM 600
Db      541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPM 600
Qy      601 CAITQGVQFIEKTCPGATSRTANFLILYGFIPALVILGIVIFKVRDYILIS 652
Db      601 CAITQGVQFIEKTCPGATSRTANFLILYGFIPALVILGIVIFKVRDYILIS 652

```

# RESULT 3

US-09-837-992-3

; Sequence 3, Application US/09837992

; Patent No. US20020081687A1

; GENERAL INFORMATION:

; APPLICANT: Tian, Hui

; APPLICANT: Schultz, Joshua

; APPLICANT: Shan, Bei

; APPLICANT: Tularik Inc.

; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions

; TITLE OF INVENTION: and Methods of Use

; FILE REFERENCE: 018781-006020US

; CURRENT APPLICATION NUMBER: US/09/837,992

; CURRENT FILING DATE: 2001-04-18

; PRIOR APPLICATION NUMBER: US 60/198,465

; PRIOR FILING DATE: 2000-04-18

; PRIOR APPLICATION NUMBER: US 60/204,234

; PRIOR FILING DATE: 2000-05-15

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3

; LENGTH: 651  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: human sitosterolemia susceptibility gene (SSG)  
; OTHER INFORMATION: amino acid sequence  
US-09-837-992-3

Query Match 81.5%; Score 2744.5; DB 9; Length 651;  
Best Local Similarity 80.2%; Pred. No. 1.4e-258;  
Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;

```
Qy      1 MGELPFLSPEGARGPHINRGSLSLSEQSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS 60
      ||:| |:| |:| :||| ||| | | |||:| | |||:| | |||:| |
Db      1 MGDLSLTPGGSMGLQVNRGSQSSLEGAPATAPEP-HSLGILHASYSVSHRVRPWWDITS 59

Qy     61 CQQKWDRQILKDVSLYIESGQIMCILGSSSGSKTTLDDAISGRLRRTGTLEGEVFNNGCE 120
      |:|:| ||||| ||||:||||| ||||| ||||:| | | |||:| |
Db     60 CRQQWTRQILKDVSLYVESGQIMCILGSSSGSKTTLDDAMSGRLGRAGTFLGEVYVNGRA 119

Qy    121 LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNKKVEAVMTLSLSH 180
      |||:||||| ||||| ||||:| | : : ||||| |||||
Db    120 LRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRGNPGSFQKKVEAVMAELSLSH 179

Qy    181 VADQMIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240
      |||:| |:|:| |||:||||| ||||:| | ||||| ||||:| | |||
Db    180 VADRLIGNYSLGGISTGERRRVSIQAQLLQDPKVMMLFDEPTTGLDCMTANQIVVLLVELA 239

Qy    241 RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNGCYPCPEHSNPF 300
      ||:| |:|:| ||||| ||||:| |:|:| ||| |||:| ||||| |||||
Db    240 RRNRIVVLTIHQPRSELFQLFQDKIAILSFGEILFCGTPEEMLDFFNDCGYPCPEHSNPF 299

Qy    301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPVFKTK 360
      ||||| ||||:| ||||| ||||:| |:|:| | | |:| | | |:| |||||
Db    300 FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMPVFKTK 359

Qy    361 DPPGMFGKLGVLRLRRVTRNLNRNKQAVIMRLVQNLIMGLFLIFYLLRVQNTLKGAVQDR 420
      | |:| | ||||| ||||:| | ||| |:| ||||| |:|:| |:|:| | |||:| |
Db    360 DSPGVFSKLGVLRLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDR 419

Qy    421 VGLLYQLVGATPYTGMLNAVNLFPMRLRAVSDQESQDGLYHKWQMLLAYVLHVLFPFSVIAT 480
      ||||| ||||| ||||| ||||:| ||||| ||||:| | ||||| ||||:| |
Db    420 VGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLFPFSVVAT 479

Qy    481 VIFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI 540
      :||| | ||||| ||||:| ||||| ||||:| ||||| ||||:| |||||
Db    480 MIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSVALLSI 539

Qy    541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHMP 600
      |:|:| |:|:| ||||| |||:| ||||| ||||| ||||| ||| |:| :| |
Db    540 AGVLVSGSGLRNIQEMPIPFKIISYFTFQKYCCEILVVNEFYGLNFTCGGSNVSVTTPNM 599

Qy    601 CAITQGVQFIEKTCPGATSRETFANFLILYGFIPALVILGIVIFKVRDYLIISR 652
      || |||:||||| ||||| ||||| ||||:| |:|:| |||
Db    600 CAFTQGIQFIEKTCPGATSRETFMNFILYSFIPALVILGIVVFKIRDHLISR 651
```



RESULT 4

US-09-989-981A-6

; Sequence 6, Application US/09989981A

; Publication No. US20030049730A1

; GENERAL INFORMATION:

; APPLICANT: Hobbs, Helen H.

; APPLICANT: Shan, Bei

; APPLICANT: Barnes, Robert

; APPLICANT: Tian, Hui

; APPLICANT: Tularik Inc.

; APPLICANT: Board of Regents, The University of Texas System

; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use

; FILE REFERENCE: 018781-007320US

; CURRENT APPLICATION NUMBER: US/09/989,981A

; CURRENT FILING DATE: 2002-07-23

; PRIOR APPLICATION NUMBER: US 60/252,235

; PRIOR FILING DATE: 2000-11-20

; PRIOR APPLICATION NUMBER: US 60/253,645

; PRIOR FILING DATE: 2000-11-28

; NUMBER OF SEQ ID NOS: 13

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 6

; LENGTH: 651

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: human ABCG5 (hABCG5)

US-09-989-981A-6

Query Match 81.5%; Score 2744.5; DB 10; Length 651;

Best Local Similarity 80.2%; Pred. No. 1.4e-258;

Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;

```

Qy      1 MGELPFLSPEGARGPHINRGSLSSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS 60
      ||:| |:| |:| :||| ||| | | |||:| | |||:| | |||:| |
Db      1 MGDLSLTPGGSMGLQVNRGSQSLEGAPATAPEP-HSLGILHASYSVSHRVRPWWDITS 59

Qy      61 CQQKWDQRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVC 120
      |:|:| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      60 CRQQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRA 119

Qy      121 LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTLSLSH 180
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      120 LRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSH 179

Qy      181 VADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAE 240
      |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      180 VADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMMLFDEPTTGLDCMTANQIVVLLVELA 239

Qy      241 RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCGYPCPEHSNPF 300
      ||:|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      240 RRNRIVVLTIHQPRSELFQLEFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPF 299

Qy      301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPVFKTK 360
      |||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      300 FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSAICHKTLKNIERMKHLKTLPMVPFKTK 359

```

Qy	361	DPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNTLKGAQDR	420
		:       :            :     : ::  :     :	
Db	360	DSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLVRVSNVLKGAIQDR	419
Qy	421	VGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLPPFSVIAT	480
		:         :       :	
Db	420	VGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLPPFSVAT	479
Qy	481	VIFSSVCYWTLGLYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIVNSIVALLSI	540
		:     :           :	
Db	480	MIFSSVCYWTLGLHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIVNSVALLSI	539
Qy	541	SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPM	600
		: : : : :          :                         :	
Db	540	AGVLVSGSGLRNIQEMPIPFKII SYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPM	599
Qy	601	CAITQGVQFIEKTCPGATSRFTANFLILYGFIPALVILGIVIFKVRDYLIISR	652
		:                 : : :	
Db	600	CAFTQGIQFIEKTCPGATSRFTMNFLILYSFIPALVILGIVVFKIRDHLISR	651

```

Query Match      81.5%; Score 2744.5; DB 14; Length 651;
Best Local Similarity 80.2%; Pred. No. 1.4e-258;
Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;

Qy      1 MGELPFLSPEGARGPHINRGSLSSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS 60
      ||:| |:| |:| :||| ||| | | |||:| | |||:| |
Db      1 MGDLSLSTPGGSMGLQVNRGSQSSLEGAPATAPEP-HSLGILHASYSVSHRVRPWWDITS 59

Qy      61 CQQKWDRQILKDVSLYIESGQIMCILGSSSGSKTTLLDAISGRLRRTGTLEGEVFNNGCE 120
      |:|:| |||||:|||||:|||||:|||| | || ||:|
Db      60 CRQQWTRQILKDVSLYVESGQIMCILGSSSGSKTTLLDAMSGRLGRAGTFLGEVYVNGRA 119

Qy      121 LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTLSLSH 180
      |||:||||| ||||| |||||:|:| : : ||||| |||||
Db      120 LRREQFQDCFSYVLQSDTLSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSH 179

Qy      181 VADQMIGSYNFGGSISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240

```

Db	180	VADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVLLVELA	239
Qy	241	RRDRIVIVTIHQPRSELFQHFDKIAILTYGELVFCGTPEEMLGFFNCGYPCPEHSNPFD	300
Db	240	RRNRIVVLTIHQPRSELFQLFDKIAILSFGEIIFCGTPAEMLDFFNDCGYPCPEHSNPFD	299
Qy	301	FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPFKTK	360
Db	300	FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMPFKTK	359
Qy	361	DPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNTLKGAVQDR	420
Db	360	DSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDR	419
Qy	421	VGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLPPFSVIAT	480
Db	420	VGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLPPFSVVAT	479
Qy	481	VIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI	540
Db	480	MIFSSVCYWTGLHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSVVALLSI	539
Qy	541	SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPM	600
Db	540	AGVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPM	599
Qy	601	CAITQGVQFIEKTCPGATSRTANFLILYGFIPALVILGIVIFKVRDY LISR	652
Db	600	CAFTQGIQFIEKTCPGATSRTMNFILYSFIPALVILGIVVFKIRDHLISR	651

RESULT 6

US-10-104-047-2795  
; Sequence 2795, Application US/10104047  
; Publication No. US20030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2795  
; LENGTH: 256  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-104-047-2795

Query Match 34.9%; Score 1177; DB 15; Length 256;  
Best Local Similarity 85.5%; Pred. No. 3.5e-106;  
Matches 219; Conservative 23; Mismatches 14; Indels 0; Gaps 0;

Qy 397 MGLFLIFYLLRVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQD 456

```

      |||||:|::|||::| ||||:||||||| |||||:|||||||:|||||||
Db      1 MGLFLLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFPVLRVSDQESQD 60

Qy      457 GLYHKWQMLLAYVLHVL PFSVIATVIFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGEF 516
      ||| ||||:||| |||||:|::|||:|||||||:|||||||:|||||||
Db      61 GLYQKWQMLLAYALHVL PFSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEF 120

Qy      517 LTLVLLGIVQNPNI VNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEIL 576
      |||||:|||||:|::|||:|||||:||||||| ||: ||||| |||
Db      121 LTLVLLGIVQNPNI VNSVALLSIAGVLVGSGLRNIQEMPIPFKIISYFTFQKYCSEIL 180

Qy      577 VVNEFYGLNFTCGGSNTSMLNHMPMAITQGVQFIEKTCPGATS RFTANFLILYGFIPALV 636
      |||||:||||| || |: :||| ||:||||| ||||| |||||
Db      181 VVNEFYGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPGATS RFTMNFILYSFIPALV 240

Qy      637 ILGIVIFKVRDY LISR 652
      |||||:|::|||:||||
Db      241 ILGIVVFKIRDHLISR 256

```

# RESULT 7

US-09-989-981A-4

```

; Sequence 4, Application US/09989981A
; Publication No. US20030049730A1
; GENERAL INFORMATION:
; APPLICANT: Hobbs, Helen H.
; APPLICANT: Shan, Bei
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 672
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: mouse ABCG8 (mABCG8)
US-09-989-981A-4

```

```

Query Match          20.8%; Score 701.5; DB 10; Length 672;
Best Local Similarity 29.1%; Pred. No. 4.2e-59;
Matches 194; Conservative 131; Mismatches 245; Indels 97; Gaps 19;

```

```

Qy      27 QGSVTGTEARHSLGVLHVSYS-----VSNRVGPW-----WNIKS 60
      | |: :|: :| |: :|| :::| || | |
Db      24 QDSLFSSES DNS---LYFTYSQSNTLEVRDLTYQVDIASQV-PWFEQLAQFKIPWRSHS 79

```



```
; CURRENT APPLICATION NUMBER: US/09/961,086
; CURRENT FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/073,763
; PRIOR FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US99/02577
; PRIOR FILING DATE: 1999-02-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-961-086-1
```

```
Query Match          20.6%; Score 693.5; DB 10; Length 655;
Best Local Similarity 29.0%; Pred. No. 2.4e-58;
Matches 181; Conservative 142; Mismatches 246; Indels 55; Gaps 16;
```

```
Qy      25 LEQGSVTGTEARHS-----LGVLHVSYSVSNRVGPWWNIKSCQQKWDRQILKDV 73
      : ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      12 VSQGNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLGSG----FLPCRKPVEKEILSNI 67

Qy      74 SLYIESGQIMCILGSSSGSKTTLDDAISGRLLRRTGTLEGEVFNVCCELRRDQFQDCFSYV 133
      : :: | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      68 NGIMKPG-LNAILGPTGGGKSSLLDVLAAARKDPSG-LSGDVLINGAP-RPANFKCNSGYV 124

Qy     134 LQSDVFLSSSLTVRETLRYTAMLALCRSSADF-YNKKVEAVMTELSLSHVADQMIGSYNFG 192
      : | | | : : | | | | | | | | | | | | | | | | | | | | | | | | |
Db     125 VQDDVVMGTLTVRENLFQSAALRLATMTNHEKNERINRVIQELGLDKVADSKVGTQFIR 184

Qy     193 GISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLLAELARRDRIVIVTIHQ 252
      |: | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     185 GVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKRMSKQGRTIIFSIIHQ 244

Qy     253 PRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPFDFYMDLTSVDQTQ- 311
      || : |: | | : : | | | | | | | | | | | | | | | | | | | | | |
Db     245 PRYSIFKLFDLTLASGRMLFMFHGPAQEALGYFESAGYHCEAYNNPADFFLDIINGDSTA 304

Qy     312 -SREREIETYKRVQMLECAFKESDIYHKI-----LENIERARYLKT 351
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     305 VALNRE-EDFKATEIIEPSKQDKPLIEKLAIEIVNSSFYKETKAELHQLSGGEKKKKITV 363

Qy     352 LPMVPFVKTKDPPGMFGLGVLLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNN 411
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     364 FKEISYTT---SFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 419

Qy     412 TLKGAQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYVL- 470
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     420 ST--GIQNRAGVLFFLTNNQCFSS-VSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLS 476

Qy     471 HVLPPFSVIATVIFSSVCYWTGLYPEVARFGYFSAALLAPHLIGEFLLTVLLGIVQNPNI 530
      : | | : : : | | | | | | | | | | | | | | | | | | | | | | |
Db     477 DLLPMTMLPSIIFTICIVYFMLGLKPKADAFFVMMFTLM---MVAYSASSMALAIAAGQSV 533

Qy     531 VNSIVALLSIS--GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTC 588
      |: | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     534 VSVATLLMTICFVFMFIFSGLLVNLTTIASWLSWLQYFSIPRYGFTALQHNEFLGQNF-C 592
```

Qy 589 GGSNTSMLNHPMCAITQGVQFIEK 612  
| | : | | : : |  
Db 593 PGLNATGNNPCNYATCTGEEYLVK 616

RESULT 9

US-10-405-806-13

; Sequence 13, Application US/10405806  
; Publication No. US20030232362A1  
; GENERAL INFORMATION:  
; APPLICANT: KOMATANI, HIDEYA  
; APPLICANT: HARA, YOSHIKAZU  
; APPLICANT: KOTANI, HIDEHITO  
; APPLICANT: NAKAGAWA, RINAKO  
; TITLE OF INVENTION: DRUG RESISTANT GENE AND USE THEREOF  
; FILE REFERENCE: 234985USOCONT  
; CURRENT APPLICATION NUMBER: US/10/405,806  
; CURRENT FILING DATE: 2003-04-03  
; PRIOR APPLICATION NUMBER: PCT/JP01/08112  
; PRIOR FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: JP2000-303441  
; PRIOR FILING DATE: 2000-10-03  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 13  
; LENGTH: 655  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: ABCG2 482Tmutant sequence  
US-10-405-806-13

Query Match 20.6%; Score 693.5; DB 15; Length 655;  
Best Local Similarity 29.0%; Pred. No. 2.4e-58;  
Matches 181; Conservative 142; Mismatches 246; Indels 55; Gaps 16;

Qy 25 LEQGSVTGTGTEARHS-----LGVLHVSYSVSNRVGPWWNIKSCQQKWDRQILKDV 73  
: ||: | | | | : : | | : | | : : || : :  
Db 12 VSQGNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLKSG----FLPCRKPVEKEILSNI 67  
  
Qy 74 SLYIESGQIMCILGSSGSGKTTLLDAISGRRLRRTGTLEGEVVFVNGCELRRDQFQDCFSYV 133  
: : : | : ||| : | ||: ||| : : | : | | : | | | : ||  
Db 68 NGIMKPG-LNAILGPTGGGKSSLLDVLAARKDPGSG-LSGDVLINGAP-RPANFKCNSGYV 124  
  
Qy 134 LQSDVFLSSLTVRETTRYTAMLALCRSSADF-YNKKVEAVMTELSLSHVADQMIGSYNFG 192  
: | || : : |||| | : : | : : | : | | | || : | :  
Db 125 VQDDVVMGTLTVRENLQFSAALRLATMTNHEKNERINRVIQELGLDKVADSKVGTQFIR 184  
  
Qy 193 GISSGERRRVSIAAQLLDQPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQ 252  
| : | ||| : | | : | : || : |||| ||| : ||| : : : | : | : |||  
Db 185 GVSggerKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKRMSKQGRITIFSIIHQ 244  
  
Qy 253 PRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCGYPCPEHSNPFDFYMDLTSVDTQ- 311  
|| : | : | : | | : | | : | | : | | : | : | : | : | :  
Db 245 PRYSIFKLFDSTLLASGRLMFHGPAQEALGYFESAGYHCEAYNNPADFFLDIINGDSTA 304

Qy	312	SRREIETYKRVQMLECAFKESDIYHKI-----LENIERARYLKT	351
		:      :  ::  : : :   :	
Db	305	VALNRE-EDFKATEIIEPSKQDKPLIEKLAEIYVNSSFYKETKAELHQLSGGEKKKKITV	363
Qy	352	LPMVPFKTKDPPGMEFGKLGVLRRVTRNLMRNKQAVIMRLVQNLMGLFLIFYLLRVQNN	411
		: :   :  : :  :  :        :: : ::   : ::  :	
Db	364	FKEISYTT----SFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND	419
Qy	412	TLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYVL-	470
		: : :   : :   : : ::      : : :         :	
Db	420	ST--GIQNRAGVLFFLTNTNQCFSS-VSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLS	476
Qy	471	HVLPFSVIATVIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFRTLVLVGIVQNPNI	530
		:   :: : :  : :  :      :   : : : : :     : :	
Db	477	DLLPMTMLPSIIFTCIVYFMLGLKPKADAFFVMMFTLM---MVAYSASSMALAIAAGQSV	533
Qy	531	VNSIVALLSIS--GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCAILVVNEFYGLNFTC	588
		: : :  : :     :  : :       : :	
Db	534	VSVATLLMTICFVFMFIFSGLLVNLTTIASWLSWLQYFSIPRYGFTALQHNEFLGQNF-C	592
Qy	589	GGNTSMLNHPMCAITQGVQFIEK	612
		:       : :	
Db	593	PGLNATGNNPCNYATCTGEEYLVK	616

US-09-981-353-35

; Sequence 35, Application US/09981353

; Patent No. US20020160382A1

; GENERAL INFORMATION:

; APPLICANT: Lasek, Amy W.

; APPLICANT: Jones, David A.

1; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER

; FILE REFERENCE: PA-0038 US

; CURRENT APPLICATION NUMBER: US/09/981,353

; CURRENT FILING DATE: 2001-10-11

; NUMBER OF SEQ ID NOS: 194

```
; SOFTWARE: PERL Program
```

; SEQ ID NO 35

; LENGTH: 655

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

```
; NAME/KEY: misc feature
```

OTHER INFORMATION: Incyte ID No. US20020160382A1 5517972CD1

US-09-981-353-35

Query Match 20.5%; Score 691.5; DB 9; Length 655;

Best Local Similarity 29.0%; Pred. No. 3.8e-58;

Matches 181; Conservative 141; Mismatches 247; Indels 55; Gaps 16;

Qy 25 LEQGSVTGTEARHS-----LGVLHVSYSVSNRVGPWWNIKSCQQKWDRQILKDV 73

$$: || : | | | | | | : : | | : | | : : : || | : :$$
[illegible]

```

:  : | : || : || : || : || : || : | : | | : | : || | : |

```



Db 68 NGIMKPG-LNAILGPTGGGKSSLLDVLAAKDPGSG-LSGDVLLINGAP-RPANFKCNSGYV 124

Qy 134 LQSDVFLSSSLTVRETLRYTAMLALCRSSADF-YNKKVEAVMTELSLSHVADQMIGSYNFG 192  
 :| || : :| || | : : : :| :| | || :| :

Db 125 VQDDVVMGTLTVRENLQFSAAALRLATTMTNHEKNERINRVIQELGLDKVADSKVGTQFIR 184

Qy 193 GISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQ 252  
 |:| |||:| || :| :| :| :| || ||| :| :| :| :| :

Db 185 GVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVLLLLKRMASKQGRTIIFSIIHQ 244

Qy 253 PRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCCGYPCPEHSNPFDFYMDLTSVDTQ- 311  
 || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :

Db 245 PRYSIFKLFDSLTLASGRIMFHGPAQEALGYFESAGYHCEAYNNPADFFLDIINGDSTA 304

Qy 312 -SREREIETYKRVQMLECAFKESDIYHKI-----LENIERARYLKT 351  
 : || | :| :| :| :| :| :| :| :| :| :| :| :| :| :

Db 305 VALNRE-EDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETKAELHQLSGGEKKKKITV 363

Qy 352 LPMVPFKTKDPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNN 411  
 : : | :| :| :| :| :| :| :| :| :| :| :| :| :| :

Db 364 FKEISYTT----SFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 419

Qy 412 TLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYVL- 470  
 : :| :| :| :| :| :| :| :| :| :| :| :| :| :| :

Db 420 ST--GIQNRAGVLFFLTNQCFS--VSAVELFVVEKKLFIHEYISGYRVSYSYFLGKLLS 476

Qy 471 HVLFPFSVIATVIFSSVCYWTGLYPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPNI 530  
 :|| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :

Db 477 DLLPMRMLPSIIFTCIVYFMLGLKPKADAFFVMMFTLM---MVAYSASSMALAIAAGQSV 533

Qy 531 VNSIVALLSIS--GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTC 588  
 |:| :| :| :| :| :| :| :| :| :| :| :| :| :| :

Db 534 VSVATLLMTICFVFMIFSGLLVNLTTIASWLSWLQYFSIPRYGFTALQHNEFLGQNF-C 592

Qy 589 GGSNTSMLNHPMCAITQGVQFIEK 612  
 || : | | :| :| :

Db 593 PGLNATGNNPCNYATCTGEEYLVK 616

RESULT 11

US-10-120-687-61

; Sequence 61, Application US/10120687

; Publication No. US20030082155A1

; GENERAL INFORMATION:

; APPLICANT: Massachusetts General Hospital

; TITLE OF INVENTION: Stem Cells of the Islets of Langerhans and Their Use in Treating Diabetes

; TITLE OF INVENTION: Mellitus

; FILE REFERENCE: 3284/1235B

; CURRENT APPLICATION NUMBER: US/10/120,687

; CURRENT FILING DATE: 2002-04-11

; PRIOR APPLICATION NUMBER: US60/169082

; PRIOR FILING DATE: 1999-12-06

; PRIOR APPLICATION NUMBER: US 09/963,875

; PRIOR FILING DATE: 2001-09-25

; PRIOR APPLICATION NUMBER: US 60/215109

; PRIOR FILING DATE: 2000-06-28

```
; PRIOR APPLICATION NUMBER: US 60/238880
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 09/731261
; PRIOR FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-120-687-61
```

```
Query Match          20.5%;  Score 691.5;  DB 14;  Length 655;
Best Local Similarity 29.0%;  Pred. No. 3.8e-58;
Matches 181;  Conservative 141;  Mismatches 247;  Indels 55;  Gaps 16;
```

```
Qy      25 LEQGSVTGTTEARHS-----LGVLHVSYSVSNRVGPWWNIKSCQQKWDRQILKDV 73
      : ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      12 VSQGNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLGSG----FLPCRKPVEKEILSNI 67

Qy      74 SLYIESGQIMCILGSSSGSGKTTLLDAISGRLRRTGTLEGEVFNVCCELRRDQFQDCFSYV 133
      : :: | : ||| : | ||: ||| : : | : | | | : | | | : | | | |
Db      68 NGIMKPG-LNAILGPTGGGKSSLLDVLAAKDPGSG-LSGDVLINGAP-RPANFKCNSGYV 124

Qy     134 LQSDVFLSSLTVRETLRYTAMLALCRSSADF-YNKKVEAVMTELSLSHVADQMIGSYNFG 192
      : | || : : |||| |::| | | : : | ::: | : || | || | : | :
Db     125 VQDDVVMGTLTVRENLFQSAALRLATTMTNHEKNERINRVIQELGLDKVADSKVGTQFIR 184

Qy     193 GISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQ 252
      | : | ||| : | | : | : || : : |||| |||| | | | : || | : : : | : | : ||
Db     185 GVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVLLLLKRMSKQGRTIIFSIIHQ 244

Qy     253 PRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNFCGYPCPEHSNPFDFYMDLTSVDTQ- 311
      || : | : | | : : | | | : | | : | | | : | | | : | | | : | : | :
Db     245 PRYSIFKLFDSLTLASGRILMFHGPQAEALGYFESAGYHCEAYNNPADFFLDIINGDSTA 304

Qy     312 -SREREIETYKRVQMLECAFKESDIYHKI-----LENIERARYLKT 351
      : | | | : | : : | : : : | : | | | | | | | | | | | | | | | |
Db     305 VALNRE-EDFKATEIIEPSKQDKPLIEKLAEIYVNSSFYKETKAELHQLSGGEKKKKITV 363

Qy     352 LPMVPFKTKDPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNN 411
      : : | : : | : : | : | : | | | : : | : : | : : | : : | : :
Db     364 FKEISYTT----SFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 419

Qy     412 TLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYVL- 470
      : : : | | : | : : : : : | | | | | | | | | | | | | | | |
Db     420 ST--GIQNRAGVLFFLTNQCFS--VSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLS 476

Qy     471 HVLPFVSVIATVIFSSVCYWTGLYPEVAREFGYFSAALLAPHLIGEFLLTVLLGIVQNPNI 530
      : || : : : | : | : | | | : | : : : : : : : | | : :
Db     477 DLLPMRMLPSIIFTICIVYFMLGLKPKADAFFVMMFTLM---MVAYSASSMALAIAAGQSV 533

Qy     531 VNSIVALLSIS--GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTC 588
      | : | : | : : | | : | : : | | | : | | | | | | | | | |
Db     534 VSVATLLMTICFVFMIFSGLLVNLTTIASWLSWLQYFSIPRYGFTALQHNEFLGQNF-C 592

Qy     589 GGSNTSMLNHPMCAITQGVQFIEK 612
```

Db 593 PGLNATGNNPCNYATCTGEEYLVK 616

RESULT 12

US-10-405-806-2

; Sequence 2, Application US/10405806

; Publication No. US20030232362A1

; GENERAL INFORMATION:

; APPLICANT: KOMATANI, HIDEYA

; APPLICANT: HARA, YOSHIKAZU

; APPLICANT: KOTANI, HIDEHITO

; APPLICANT: NAKAGAWA, RINAKO

; TITLE OF INVENTION: DRUG RESISTANT GENE AND USE THEREOF

; FILE REFERENCE: 234985US0CONT

; CURRENT APPLICATION NUMBER: US/10/405,806

; CURRENT FILING DATE: 2003-04-03

; PRIOR APPLICATION NUMBER: PCT/JP01/08112

; PRIOR FILING DATE: 2001-09-18

; PRIOR APPLICATION NUMBER: JP2000-303441

; PRIOR FILING DATE: 2000-10-03

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 2

; LENGTH: 655

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-405-806-2

Query Match 20.5%; Score 691.5; DB 15; Length 655;

Best Local Similarity 29.0%; Pred. No. 3.8e-58;

Matches 181; Conservative 141; Mismatches 247; Indels 55; Gaps 16;

```
Qy      25 LEQGSVTGTEARHS-----LGVLHVSYSVSNRVGPWWNIKSCQQKWDRQILKDV 73
      : ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      12 VSQGNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLKSG----FLPCRKPVEKEILSNI 67

Qy      74 SLYIESGQIMCILGSSSGSGKTTLLDAISGRLRRTGTLEGEVFNVCGLRRDQFQDCFSYV 133
      : :: | : ||| :| ||::||| :: | :| | | :| :| | | :| |
Db      68 NGIMKPG-LNAILGPTGGGKSSLLDVLAARKDPSG-LSGDVLLINGAP-RPANFKCNSGYV 124

Qy     134 LQSDVFLSSLTVRETLRYTAMLALCRSSADF-YNKKVEAVMTELSLSHVADQMIGSYNFG 192
      :| || : :||| | | | | | | | | | | | | | | | | | | | | | |
Db     125 VQDDVVMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIQELGLDKVADSKVGTQFIR 184

Qy     193 GISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLLAELARRDRIVIVTIHQ 252
      |:| |||:| || :|: || :: ||||| ||| :||| :::: | :| :|||
Db     185 GVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVLLLLKRMSKQGRTIIFSIIHQ 244

Qy     253 PRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNGGYPCPEHSNPFDFYMDLTSVDTQ- 311
      || :|: || : :| | | | | :| ||| : | | | :|| | | :| : |
Db     245 PRYSIFKLFDSLTLASGRIMFHGPAQEALGYFESAGYHCEAYNNPADFFLDIINGDSTA 304

Qy     312 -SREREIETYKRVQMLECAFKESDIYHKI-----LENIERARYLKT 351
      : || | :| ::| : :: : | : | | | | | | | | | | | | | | | |
Db     305 VALNRE-EDFKATEIIEPSKQDKPLIEKLAETVNSSFYKETKAEHLQLSGGEKKKKITV 363
```

Qy	352	LPMVPFKTKDPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLMGLFLIFYLLRVQNN	411
		: :   :   : :   :   : : : : : : : : : : : :	
Db	364	FKEISYTT----SFCHQLRWVSKRSFKNLLGNPQASIAQIIIVTVVLGLVIGAIYFGLKND	419
Qy	412	TLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVL-	470
		: : :   : :   : : : : : : : : : : :   :   :   :	
Db	420	ST--GIQNRAGVLFFLTNTQCFSS-VSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLS	476
Qy	471	HVLPFSVIATVIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFITLVLLGIVQNPNI	530
		:   :	
Db	477	DLLPMRMLPSIIFTCIVYFMLGLKPKADAFFVMMFTLM---MVAYSASSMALAIAAGQSV	533
Qy	531	VNSIVALLSIS--GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTC	588
		: :	
Db	534	VSVATLLMTICFVMMIFSGLLVNLTTIASWLSWLQYFSIPRYGFTALQHNEFLGQNF-C	592
Qy	589	GGSENTSMLNHPMCAITQGVQFIEK	612
		: :   : : : : : : : : : : : : : : : : : :	
Db	593	PGLNATGNNPCNYATCTGEEYLVK	616

## RESULT 13

```

US-09-866-866A-10
; Sequence 10, Application US/09866866A
; Patent No. US20020102244A1
; GENERAL INFORMATION:
; APPLICANT: Sorrentino, Brian
; APPLICANT: Schuetz, John
; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
; FILE REFERENCE: 1340-1-021CIP2
; CURRENT APPLICATION NUMBER: US/09/866,866A
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: 09/584,586
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: PCT/US99/11825
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: 60/086,988
; PRIOR FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-866-866A-10

```

Query Match 20.5%; Score 689.5; DB 9; Length 655;  
Best Local Similarity 29.0%; Pred. No. 6e-58;  
Matches 181; Conservative 141; Mismatches 247; Indels 55; Gaps 16;

```

Qy      25 LEQGSVTGTEARHS-----LGVLHVSYSVSNRVGPWWNIKSCQQKWDRQILKDV 73
      : ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      12 VSQGNTNGFPATVSNDLKAFTEGAVLSFHNICYRVKLKSG----FLPCRKPVEKEILSNI 67

Qy      74 SLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVCELRRDQFQDCFSYV 133
      :  :: | : ||| : | ||::||| :: | : | | | : | | | | | | | | | |
Db      68 NGIMKPG-LNAILGPTGGGKSSLLDVLAARKDPSG-LSGDVLINGAP-RPANFKCNSGYV 124

```

Qy 134 LQSDVFLSSLTVRETLRYTAMLALCRSSADF-YNKKVEAVMTELSLSHVADQMIGSYNFG 192  
 :| || : :||| |::| | | : : |:: | : || | || | :|  
 Db 125 VQDDVVMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIEELGLDKVADSKVGTQFIR 184

Qy 193 GISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQ 252  
 |:| |||:| || :|: || :: ||||| || | :||| ::: | :| :||  
 Db 185 GVSGGERKRTSIGMELITDPSILSLDEPTTGLDSSTANAVLLLLKMSKQGRTIIFSIIHQ 244

Qy 253 PRSELFQHFDKIAILTYGELVFCGTPEEMLGFFNCCGYPCPEHSNPFDFYMDLTSVDTQ- 311  
 || :|: || : :| | | :| | :| ||:| : || | :|| ||::|: : |:  
 Db 245 PRYSIFKLFDLSLTLLASGRIMFHGPAQEALGYFESAGYHCEAYNNPADFFLDIINGDSTA 304

Qy 312 -SREREIETYKRVQMLECAFKESEDYHKKI-----LENIERARYLKT 351  
 : || | :| :|| : : : : | : | : : :  
 Db 305 VALNRE-EDFKATEIIEPSKQDKPLIEKLAEIYVNSSFYKETKAELHQLSGGEKKKKITV 363

Qy 352 LPMVPFKTKDPPGMFEGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNN 411  
 : : | :| : :| :||: | || | ::: :::|| : :||:  
 Db 364 FKEISYTT----SFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 419

Qy 412 TLKGAQVQDRVGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVL- 470  
 : :|| |::| : : : :|| || : : : | | | :|  
 Db 420 ST--GIQNRAGVLFFLTNQCFFS-VSAVELFVVEKKLFIHEYISGYRVSYSYFLGKLLS 476

Qy 471 HVLPPFSVIATVIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPNI 530  
 :|| : :||: : | : || | : | : : : : : : : :  
 Db 477 DLLPMRMLPSIIFTCIVYFMLGLKPKADAFFVMMFTLM--MVAYSASSMALAIAAGQSV 533

Qy 531 VNSIVALLSIS--GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCILVNEFYGLNFTC 588  
 |: :|| :::| || : | : : | | ||: :| | || | || |  
 Db 534 VSVATLLMTICFVFMIFSGLLVNLTTIASWLSWLQYFSIPRYGFTALQHNEFLGQNF-C 592

Qy 589 GGSNTSMLNHPMCAITQGVQFIEK 612  
 | | : | | | : : |  
 Db 593 PGLNATGNNPCNYATCTGEEYLVK 616

RESULT 14

US-10-090-455-5

; Sequence 5, Application US/10090455

; Publication No. US20030027259A1

; GENERAL INFORMATION:

; APPLICANT: Chen, Hongyun

; APPLICANT: Le Bihan, Stephane

; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF

; FILE REFERENCE: 100103.406

; CURRENT APPLICATION NUMBER: US/10/090,455

; CURRENT FILING DATE: 2002-03-01

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 5

; LENGTH: 655

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-090-455-5

Query Match 20.5%; Score 689.5; DB 14; Length 655;  
Best Local Similarity 29.0%; Pred. No. 6e-58;  
Matches 181; Conservative 141; Mismatches 247; Indels 55; Gaps 16;

```
Qy      25 LEQGSVTGTEARHS-----LGVLHVSYSVSNRVGPWWNIKSCQQKWDRQILKDV 73
      : ||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      12 VSQGNTNGFPATVSNDLKAFTEGAVLSFHNICYRVKLGSG----FLPCRKPVEKEILSNI 67

Qy      74 SLYIESGQIMCILGSSSGSKTTLLDAISGRLRRTGTLEGEVFNVCCELRRDQFQDCFSYV 133
      : :: | : ||| :| ||::||| :: | : | | | :| | | | | | | | | | |
Db      68 NGIMKPG-LNAILGPTGGGKSSLLDVLAARKDPSG-LSGDVLINGAP-RPANFKCNSGYV 124

Qy     134 LQSDVFLSSLTVRETLRYTAMLALCRSSADF-YNKKVEAVMTELSLSHVADQMIGSYNFG 192
      :| || : :||| | |::| | | : : |::: | : || | ||| :|:
Db     125 VQDDVVMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIEELGLDKVADSKVGTQFIR 184

Qy     193 GISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLLAELARRDRIVVTIHQ 252
      |:| |||:| || :|: || :: | ||||| || | :||| :::: | :| :|||
Db     185 GVSGGERKRTSIGMELITDPSILSLDEPTTGLDSSTANAVLLLLKRMSKQGRTIIFS IHQ 244

Qy     253 PRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCGYPCPEHSNPFDFYMDLTSVDTQ- 311
      || :|: || : :| | | :| | | :| | | :| | | :| | | | :| :| :| :|
Db     245 PRYSIFKLFDSTLLASGRIMFHGPAQEALGYFESAGYHCEAYNNPADFFLDIINGDSTA 304

Qy     312 -SREREIETYKRVQMLECAFKESDIYHKI-----LENIERARYLKT 351
      : || | :| :::| : :: : | : | | | | | | | | | | | | | | | |
Db     305 VALNRE-EDFKATEIIEPSKQDKPLIEKLAIEIYVNSSFYKETKAELHQLSGGEKKKKITV 363

Qy     352 LPMVPFKTKDPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLMGLFLIFYLLRVQNN 411
      : : | : | : :| :||: | || | ::: :::|| : : : :| :
Db     364 FKEISYTT----SFCHQLRWVSKRSFKNLLGNPQASIAQIIIVTVVLGLVIGAIYFGLKND 419

Qy     412 TLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVL- 470
      : :|: | |:| : : : :|| || : : : | | | | | | | | :|
Db     420 ST--GIQNRAVLFFLTNQCFFS-VSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLS 476

Qy     471 HVLPFVSVIATVIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFLLTVLLGIVQNPNI 530
      :|| :: :||: : | : ||| | : | | : : : : : | | : :
Db     477 DLLPMRMLPSIIIFTCIVYFMLGLKPKADAFFVMFTLM--MVAYSASSMALAIAAGQSV 533

Qy     531 VNSIVALLSIS--GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTC 588
      |: | :| : :| || : | : | | | | :| | | | | | | | | |
Db     534 VSVATLLMTICFVFMIFSGLLVNLTITASWLSWLQYFSIPRYGFTALQHNEFLGQNF-C 592

Qy     589 GGSNTSMLNHPMCAITQGVQFIEK 612
      | | : | | | : : |
Db     593 PGLNATGNNPCNYATCTGEEYLVK 616
```

RESULT 15

US-09-989-981A-8

; Sequence 8, Application US/09989981A

; Publication No. US20030049730A1

; GENERAL INFORMATION:

; APPLICANT: Hobbs, Helen H.

; APPLICANT: Shan, Bei

; APPLICANT: Barnes, Robert

```
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 673
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human ABCG8 (hABCG8)
US-09-989-981A-8
```

```
Query Match          20.4%; Score 688.5; DB 10; Length 673;
Best Local Similarity 28.1%; Pred. No. 7.8e-58;
Matches 188; Conservative 125; Mismatches 233; Indels 123; Gaps 16;
```

```
Qy      37 HSLGVLHVSYSV--SNRVGPW-----WNIKSCQQKWDRQILKDVSLYIESGQIMC 84
      ::| | ::| | :::| || | || | : | :::| : ||::
Db      45 NTLEVRDLNYQVDLASQV-PWFEQLAQFKMPWTSPSCQNSCELGI-QNLSFKVRSGQMLA 102

Qy      85 ILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVCCELRRDQFQDCFSYVLQSDVFLSSLT 144
      |:|||| |: :||| |:|| |:::|| | : | :| | : | :||
Db     103 IIGSSGCGRASLLDVITGRGHGGKIKSGQIWINGQPSSPQLVRKCAHVVRQHNQLLPNLT 162

Qy     145 VRETLRYTAMLALCRS-SADFYNNKVEAVMTELSLSHVADQMIGSYNFGGSISSGERRRVS 203
      ||||| : | : | |: | :||| |: || | || :|: | :| |||||
Db     163 VRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVS 222

Qy     204 IAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQHFDDK 263
      | ||| :| :::|||||:|||| ||: :| |: ||: :|:|:::|||||:|: ||
Db     223 IGVQLLWNPGLILDEPTSGLDSTAHNLVKTLSRLAKGNRLVLISLHQPRSDIFRLFDL 282

Qy     264 IAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRV 323
      : ::| | :: | : |: | |||| :||| |||:||||:| :|||:|: | ::
Db     283 VLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKA 342

Qy     324 QMLECAFKESDIYHKILENIERARYL-----KTLPM----VPFKT 359
      | | | | :|: | | | : | :| |
Db     343 QSLAALF-----LEKVRDLDDFLWKAETKDLDCTCVESVTPLDTNCLPSPT 390

Qy     360 KDPPGMFGKLGVLRLRRVTRNLMRNKQAVIMRLVQNLMGLFLIFYLLRVQNNTLKGAVQD 419
      | || : |:|| | |: :: : :| : : | | : :: : |
Db     391 K-MPGAVQQFTTLIRRQISNDFRDLPTLLIHGAACLMSMTIGF--LYFGHGSIQLSFMD 447

Qy     420 RVGLLYQLVGATPYTGMLNAVNLFPMRLRAVSDQESQDGLYHKWQMLLAYVLHVLPPFSVIA 479
      ||: : |: :||: : ||: | :||| | :| ||
Db     448 TAALLFMIGALIPFNVILDVISKCYSERAMLYYELEDGLYTTGPYFFAKILGELPEHCAY 507
```

```

Qy      480 TVIFSSVCYWTGLGLYPEVARF-----GYFSAALLAPHLIGEFLLTLVLL 522
      :|:  ||  | | : |      :|||  :  | : |
Db      508 IIIYGMPTYWLANLRPGLQPFLHFLLVWLVVFCCRIMALAAAALLPTFHMASFFSNAL- 566

Qy      523 GIVQNPNIIVNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEF- 581
      :  : ||: |: :      :  :| :| | |: :|
Db      567 -----YNSFYLAGGEMINLSSLWTVPAWISKVSFLRWCFEGLMKIQFS 609

Qy      582 ---YGL---NFTCGGSNTSML-----NHPMCAITQGVQFIEKTCPGATSRFTANFLILY 629
      | :  | |  |  :|      ::|: ||      :||:
Db      610 RRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAI-----YLIVI 648

Qy      630 GFIPALVIL 638
      |      ::|
Db      649 GLSGGFVVL 657

```

Search completed: February 27, 2004, 07:34:03  
Job time : 31.3006 secs



OM protein - protein search, using sw model

Run on: February 27, 2004, 06:40:43 ; Search time 36.1949 Seconds  
(without alignments)  
5683.620 Million cell updates/sec

Title: US-09-989-981A-2  
Perfect score: 3369  
Sequence: 1 MGELPFLLSPEGARGPHINRG.....PALVILGIVIFKVRDY LISR 652

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertibrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result		Query			ID	Description
No.	Score	Match	Length	DB		
-----						

1	3315	98.4	652	11	Q7TSR8	Q7tsr8 mus musculu
2	717.5	21.3	672	11	Q8CIQ5	Q8ciq5 rattus norv
3	705.5	20.9	672	11	Q7TSR7	Q7tsr7 mus musculu
4	703.5	20.9	673	11	Q8R543	Q8r543 mus musculu
5	701.5	20.8	672	11	Q7TSR6	Q7tsr6 mus musculu
6	692.5	20.6	655	4	Q8IX16	Q8ix16 homo sapien
7	691.5	20.5	655	4	Q96TA8	Q96ta8 homo sapien
8	685	20.3	656	6	Q8MIB3	Q8mib3 sus scrofa
9	683.5	20.3	655	4	Q96LD6	Q96ld6 homo sapien
10	677	20.1	657	11	Q9R004	Q9r004 mus musculu
11	676	20.1	657	11	Q7TMS5	Q7tms5 mus musculu
12	659	19.6	657	11	Q80W57	Q80w57 rattus norv
13	657	19.5	657	11	Q80XF3	Q80xf3 rattus norv
14	657	19.5	657	11	Q80ST1	Q80st1 rattus norv
15	640	19.0	801	5	Q8T691	Q8t691 dictyosteli
16	636	18.9	643	5	Q7YYX5	Q7yyx5 cryptospori
17	631	18.7	672	10	Q9LI82	Q9li82 arabidopsis
18	630	18.7	725	10	Q9ZU35	Q9zu35 arabidopsis
19	630	18.7	725	10	Q9ASR9	Q9asr9 arabidopsis
20	619.5	18.4	727	10	Q9FNB5	Q9fnb5 arabidopsis
21	606.5	18.0	725	10	Q9M3D6	Q9m3d6 arabidopsis
22	603.5	17.9	668	10	Q9ARU4	Q9aru4 oryza sativ
23	602.5	17.9	541	4	Q86V64	Q86v64 homo sapien
24	597	17.7	590	10	Q9MAH4	Q9mah4 arabidopsis
25	595.5	17.7	604	5	Q8MRJ2	Q8mrj2 drosophila
26	595	17.7	609	10	Q9C8W6	Q9c8w6 arabidopsis
27	595	17.7	662	10	Q949Y4	Q949y4 arabidopsis
28	595	17.7	662	10	Q84TH5	Q84th5 arabidopsis
29	593	17.6	610	5	P90746	P90746 caenorhabdi
30	592.5	17.6	691	10	Q8RWI9	Q8rwi9 arabidopsis
31	592	17.6	679	5	Q8IS30	Q8is30 bactrocera
32	590	17.5	654	10	Q9LIW2	Q9liw2 oryza sativ
33	587	17.4	708	10	Q9M2V5	Q9m2v5 arabidopsis
34	585.5	17.4	692	10	Q7XUM2	Q7xum2 oryza sativ
35	582.5	17.3	787	10	Q8H8V7	Q8h8v7 oryza sativ
36	582	17.3	687	5	Q94960	Q94960 drosophila
37	580.5	17.2	703	10	Q8RXN0	Q8rxn0 arabidopsis
38	579.5	17.2	658	5	O16574	O16574 caenorhabdi
39	579.5	17.2	679	5	Q9BH97	Q9bh97 ceratitidis c
40	579	17.2	594	10	Q9LJC3	Q9ljc3 arabidopsis
41	578	17.2	1528	5	Q8T677	Q8t677 dictyosteli
42	577	17.1	720	10	Q9M2V7	Q9m2v7 arabidopsis
43	574	17.0	626	5	Q8T684	Q8t684 dictyosteli
44	574	17.0	626	5	Q86HQ2	Q86hq2 dictyosteli
45	573.5	17.0	608	5	Q22802	Q22802 caenorhabdi

# ALIGNMENTS

## RESULT 1

Q7TSR8

ID Q7TSR8 PRELIMINARY; PRT; 652 AA.

AC Q7TSR8;

DT 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE ATP-binding cassette sub-family G member 5.  
 GN ABCG5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=I/LnJ; TISSUE=Liver;  
 RA Wittenburg H., Lyons M.A., Li R., Churchill G.A., Carey M.C.,  
 RA Paigen B.;  
 RT "Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone  
 RT Susceptibility: Evidence from a Cross of PERA/Ei and I/Ln Inbred  
 RT Mice.";  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AY195872; AAO45093.1; -.  
 KW ATP-binding.  
 SQ SEQUENCE 652 AA; 73236 MW; 0125FB617DE296B9 CRC64;

Query Match 98.4%; Score 3315; DB 11; Length 652;  
 Best Local Similarity 98.3%; Pred. No. 1.2e-238;  
 Matches 641; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy	1	MGELPFLSPEGARGPHINRGSLSSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS	60
Db	1	MGELPFLSPEGARGPHINRGSLSSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS	60
Qy	61	CQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVC	120
Db	61	CQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVC	120
Qy	121	LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKEAVMTLSLH	180
Db	121	LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKEAVMTLSLH	180
Qy	181	VADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA	240
		:	
Db	181	VADQVIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA	240
Qy	241	RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNGYPCPEHSNPF	300
Db	241	RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNGYPCPEHSNPF	300
Qy	301	FYMDLTSVDTQSREREIETYKRVQMLECAFKESEDIYHKILENIERARYLKTLPVPFKTK	360
Db	301	FYMDLTSVDTQSREREIETYKRVQMLESAFKESEDIYHKILENIERARYLKTLPVPFKTK	360
Qy	361	DPPGMFGKLGVLRLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNLTGKAVQDR	420
Db	361	DPPGMFGKLGVLRLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNLTGKAVQDR	420
Qy	421	VGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVPFVSIAT	480
Db	421	VGLLYQFVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHALPFVSIAT	480
Qy	481	VIFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGFELTLVLLGIVQNPNIIVNSIVALLSI	540

```

Db      481 VIFSSVCYWTLGLYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI 540
Qy      541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPM 600
        |||:|||||
Db      541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGESNTTMLNHPM 600
Qy      601 CAITQGVQFIEKTCPGATSRFTANFLILYGFIPALVILGIVIFKVRDYLI SR 652
        |||:|||||
Db      601 CAITQGVFIEKTCPGATSRFTANFLILYGFIPALVILGIVIFKVRDYLI SR 652

```

# RESULT 2

## Q8CIQ5

```

ID      Q8CIQ5          PRELIMINARY;          PRT;      672 AA.
AC      Q8CIQ5;
DT      01-MAR-2003 (TrEMBLrel. 23, Created)
DT      01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Sterolin 2.
GN      ABCG8.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Sprague-Dawley;
RA      Yu H., Lu K., Lee M., Pandit B., Patel S.B.;
RT      "The rat Abcg5 and Abcg8: characterization, chromosomal assignment and
RT      genetic variation in sitosterolemic rats.";
RL      Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; AY145899; AAN64276.1; -.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0005524; F:ATP binding; IEA.
DR      GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR      GO; GO:0006810; P:transport; IEA.
DR      InterPro; IPR003439; ABC_transporter.
DR      Pfam; PF00005; ABC_tran; 1.
DR      ProDom; PD000006; ABC_transporter; 1.
DR      PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR      PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
SQ      SEQUENCE      672 AA;  75906 MW;  2FE0846E71BD9D47 CRC64;

```

```

Query Match          21.3%;  Score 717.5;  DB 11;  Length 672;
Best Local Similarity 30.7%;  Pred. No. 6.3e-45;
Matches 196;  Conservative 123;  Mismatches 245;  Indels 75;  Gaps 19;

```

```

Qy      23 SSLEQGSVTGTTEARHS-----LGVLHVSYSV--SNRVGPW-----WNIK 59
        ||| | || :|: :|          | | :|| | :|| | | :
Db      21 SSL-QDSVFSSSDNSLYFTYSGQSNLTLEVRDLTYQVDMASQV-PWFEQLAQFKLPWRSR 78
Qy      60 SCQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVC 119
        | || | | :||: | : |||: | :||: | | : || | :|| | :||: ||
Db      79 GSQDSWDLGI-RNLSFKVRSGQMLAIIGSAGCGRATLLDVITGRDHGGMKMSGQIWINQ 137
Qy      120 ELRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRS-SADFYNNKKEAVMTELSL 178
        | | :|| | | | :||| || | : | : | : | :||: || | : || |

```

Db	138 PSTPQLIQKCAVHRQQDQLLPNLTVRETLTFFIAQMRLPKTFSSQAQRDKRVEDVIAELRL	197
Qy	179 SHVADQMIGSYNFGGISSGERRRVSTAAQLLDQDPKVMMLEDEPTGLDCMTANQIVLLLLAE	238
	:   :                 :   :: :         :   :	
Db	198 RQCANTRVGNITYVRGVSGGERRRVSIGVQLLNWPGILILDEPTSGLDSFTAHLNVRTLRSR	257
Qy	239 LARRDRIVIVTIHQPRSELFQHFDKIAILTYGELVFECGTPEEMLGFFNNCGYPCPEHSNP	298
	:   :   : : :         :   :   : :   :   : :   :       :	
Db	258 LAKGNRLVLISLHQPRTSDFRLFDFLVLLMSTGPITLYLGVAQHMVQYFTSIGYPSPRYSNP	317
Qy	299 FDFYMDLTSVDTQSREIEIETYKRVMLECAFKE-----SDIYHKI-LENIERARYLKTL	352
	:       :   :   :   :   : : :       : : :   :	
Db	318 ADFYVDLTSIDRRSKEQEAVATMEKARLLAALFLEKVQGDDFLWKAEAKSLDTGYAVSQ	377
Qy	353 PMVPFKTKDP-----PGMFVKGLGVLLRRVTNRNLMRNKQAVIMRLVQNLMIGLFLIFY	404
	:   :     :   :       : : : : :   :	
Db	378 TL----TQDTNCGTAAELPMIQQTTLIRRIQISNDFRDLPTLFIHGAACLSLIIGFL	433
Qy	405 LLRVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNFMPMLRAVSDQESQDGLYHKWQM	464
	:   :     : :   : :   :   : :   :	
Db	434 YYGHADKPL--SFMDMAALLFMIGALIPFNVIDDVSKCHSERSLLYELEDGLYTAGPY	491
Qy	465 LLAYVLHVLPFSVIATVIFSSVCYWTLGLYPEVARFGYFSAALLAPHLIGEFL-----T	518
	: :             : :	
Db	492 FFAKVLGELPEHCAYVIIYGMPIYWLTNLRP-----GPFLFLHFMLLWLVVFC CRT	543
Qy	519 LVLLGIVQNPNI-VNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILV	577
	:     : : : : : : :   : :   : :   :	
Db	544 MALAASAMLP TFHMS SFCCNALYN SFYLTAGFMINLNLWIVPAWISKMS FLRWCF S GLM	603
Qy	578 VNEFYG-----LNFTCGGSN--TSM-LN-HPMCAI	603
	:       :     :       :	
Db	604 QIQFNNGHIYTTQIGNLTFSVPGDAMVTAMDLSNHPLYAI	642

### RESULT 3

```

ID      Q7TSR7      PRELIMINARY;      PRT;      672 AA.
AC      Q7TSR7;
DT      01-OCT-2003 (TrEMBLrel. 25, Created)
DT      01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      ATP-binding cassette sub-family G member 8.
GN      ABCG8.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=I/LnJ; TISSUE=Liver;
RA      Wittenburg H., Lyons M.A., Li R., Churchill G.A., Carey M.C.,
RA      Paigen B.;
RT      "Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone
RT      Susceptibility: Evidence from a Cross of PERA/Ei and I/Ln Inbred
RT      Mice.";
RL      Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.

```



08R543

Query Match 20.9%; Score 703.5; DB 11; Length 673;  
Best Local Similarity 28.7%; Pred. No. 6.9e-44;  
Matches 193; Conservative 130; Mismatches 242; Indels 107; Gaps 17;

Qy	27	QGSVTGTEARHSLGVLHVSYS-----VSNRVGPW-----WNIKS	60
		: : :   : :   : : :	
Db	25	QDSLFSSES DNS---LYFTYSGQSNLTLEVRDLTYQVDIASQV-PWFQQLAQFKIPWRSHS	80
Qy	61	CQQKWDRQILKDVSLYIESGQIMCILGSSSGSKTTLDDAISGRLLRTGTLEGEVFNVC	120
		:   : :   :     : :         : :     : : :	
Db	81	SQDSCELGI-RNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGMKSGQIWINQGP	139
Qy	121	LRRDQFQDCFSYVLQSDVFLSSSLTVRETLRYTAMLALCRS-SADFYNNKVEAVMTELSLS	179
		:   : :       :         :   :   :   : :	





RT "Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone  
 RT Susceptibility: Evidence from a Cross of PERA/Ei and I/Ln Inbred  
 RT Mice.";  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY196216; AAO45096.1; -.  
 KW ATP-binding.  
 SQ SEQUENCE 672 AA; 75867 MW; CAB720502EA8FE21 CRC64;

Query Match 20.8%; Score 701.5; DB 11; Length 672;  
 Best Local Similarity 29.1%; Pred. No. 9.8e-44;  
 Matches 194; Conservative 131; Mismatches 245; Indels 97; Gaps 19;

Qy	27	QGSVTGT	TEARHSLGVLHVSYS-----VSNRVGPW-----WN	IKS	60
		:	:   :   :   :   :   :   :   :   :   :		
Db	24	QDSLFSSES	DN---LYFTYSGQSN	TLEVRDLTYQVDIASQV-PWFEQLAQFKIPWRSHS	79
Qy	61	CQQKWDRQILKDV	SLYIESGQIMCILGSSSGSKTTLLDAISGR	LRRTGTLEGEVFVNGCE	120
		:	:   :   :   :   :   :   :   :   :   :		
Db	80	SQDSCE	LGI-RNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGMKSGQIWINGQP	138	
Qy	121	LRRDQFQDCFSYVLQSDV	FLSSSLTVRETLRYTAMLALCRS-SADFY	NKKVEAVMTELSLS	179
		:	:   :   :   :   :   :   :   :   :		
Db	139	STPQLVRKCV	AHVRQHDQLLEPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLR	198	
Qy	180	HVADQMIGSYNFGGI	SSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAEL	239	
		:	:   :   :   :   :   :   :   :   :   :		
Db	199	QCANTRVGNTYVRGV	SGGERRRVSIGVQLLWNP	GILILDEPTSGLDSFTAHLNLTLSRL	258
Qy	240	ARRDRIVIVTIHQPRSE	LQHFQDKIAILTYGELVFCGTPEEMLGFFN	NCGYPCPEHSNPF	299
		:	:   :   :   :   :   :   :   :   :   :		
Db	259	AKGNRLVLISLHQPR	SIDIFRLFDLVLMTSGTPIYLGAAQMQVQYFTSIGHPCPRYSNPA	318	
Qy	300	DFYMDLTSVDTQSRERE	IETYKRVQMLECAFKE-----SDIYH	KI-LENIERARYLKTLP	353
		:       :	:   :   :   :   :   :   :   :   :		
Db	319	DFYVDLTSIDRRSKERE	VATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTSTHTVSLT	378	
Qy	354	MVPFKTKDP-----PGM	FGLGVLLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYL	405	
		:	:   :   :   :   :   :   :   :   :		
Db	379	L----	TQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACLSLIIGF--	432	
Qy	406	LRVQNNTLKGAVQDRV	GLLYQLVGATPYTGMLNAVNLFPM	LRAVSDQESQDGLYHKWQML	465
		:	:   :   :   :   :   :   :   :   :		
Db	433	LYYGHGAKQLSFMDTA	ALLFMIGALIPFNVILDVVSKCHSERSMLY	YELEDGLYTAGPYF	492
Qy	466	LAYVLHVLPPFSVIAT	VIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFL-----TL	519	
		:	:   :   :   :   :   :   :   :   :		
Db	493	FAKILGELPEHCAYVII	YAMPYWLTLNLRPVPELF-----LL--HFLLVWL	VVFCRTM	544
Qy	520	VLLGIVQNPNI-VNSI	VALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVV	578	
		:	:   :   :   :   :   :   :   :   :		
Db	545	ALAASAMLPTFHMSS	FFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCFSGLMQ	604	
Qy	579	NEFYGL-----NFT	CGGSNTSML-----NHMCA---ITQGVQFIEKTCPGATS	RFT	622
		:	:   :   :   :   :   :   :   :   :		
Db	605	IQFNHLYTTQIGNFT	SILGDTMISAMDLNSHPLYAIYLIVIGISY-----	651	
Qy	623	ANFLILY	629		

Db                    || ||  
652 -GFLFLY 657

RESULT 6

Q8IX16

ID    Q8IX16            PRELIMINARY;            PRT;    655 AA.  
AC    Q8IX16;  
DT    01-MAR-2003 (TrEMBLrel. 23, Created)  
DT    01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT    01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE    ATP-binding cassette protein ABCG2.  
GN    ABCG2.  
OS    Homo sapiens (Human).  
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX    NCBI\_TaxID=9606;  
RN    [1]  
RP    SEQUENCE FROM N.A.  
RA    Yoshikawa M., Yabuuchi H., Ikegami Y., Ishikawa T.;  
RL    Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.  
DR    EMBL; AF463519; AAO14617.1; -.  
DR    GO; GO:0016020; C:membrane; IEA.  
DR    GO; GO:0005524; F:ATP binding; IEA.  
DR    GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
DR    GO; GO:0000166; F:nucleotide binding; IEA.  
DR    GO; GO:0006810; P:transport; IEA.  
DR    InterPro; IPR003593; AAA\_ATPase.  
DR    InterPro; IPR003439; ABC\_transporter.  
DR    InterPro; IPR006162; Ppantne\_S.  
DR    Pfam; PF00005; ABC\_tran; 1.  
DR    ProDom; PD000006; ABC\_transporter; 1.  
DR    SMART; SM00382; AAA; 1.  
DR    PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
DR    PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.  
KW    ATP-binding.  
SQ    SEQUENCE    655 AA;    72314 MW;    A8AF60B591D4C5A8 CRC64;

Query Match                    20.6%;    Score 692.5;    DB 4;    Length 655;  
Best Local Similarity    29.0%;    Pred. No. 4.4e-43;  
Matches 181;    Conservative 141;    Mismatches 247;    Indels    55;    Gaps    16;

Qy	25	LEQGSVTGTEARHS-----LGVLHVSYSVSNRVGPWWNIKSCQQKWDRQILKDV	73
		:   :	
Db	12	VSQGN TNGFPATASNDLKAFTEGAVLSFHNICYRVKLKSG----FLPCRKPVEKEILSNI	67
Qy	74	SLYIESGQIMCILGSSSGSKTTLLDAISGRLRRTGTLEGEVFNVCGLRRDQFQDCFSYV	133
		: : :   :       :       :	
Db	68	NGIMKPG-LNAILGPTGGGKSSLLDVLAARKDPSG-LSGDVLINGAP-RPANFKCNSGYV	124
Qy	134	LQSDVFLSSLTVRETLRYTAMLALCRSSADF-YNKKVEAVMTLSLSHVADQMIGSYNFG	192
		:       : :	
Db	125	VQDDVVMGTLTVRENLFSAALRLATTMTNHEKNERINRVIQELGLDKVADSKVGTQFIR	184
Qy	193	GISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVVTIHQ	252
		:	
Db	185	GVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVLLLLKRMKQGRTIIFSIIHQ	244

Qy 253 PRSELFQHFDKIAILTYGELVFCGTPEEMLGFFNCCGYPCPEHSNPFDFYMDLTSVDTQ- 311  
 || :|: || : :| | | :| | :| : || | :|:| ||::| : :| :  
 Db 245 PRYSIFKLFDSLTLASGRLMFHGPAQEALGYFESAGYHCEAYNNPADFFLDIINGDSTA 304

Qy 312 -SREREIETYKRVQMLECAFKESDIYHKI-----LENIERARYLKT 351  
 : || | :| :::| : : : | : | | : : :  
 Db 305 VALNRE-EDFKATEIIEPSKQDKPLIEKLAIEYVNSSFYKETKAELHQLSGGEKKKKITV 363

Qy 352 LPMVPFKTKDPPGMFGKLGVLRLRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNN 411  
 : : | :| : :| :||: | || | ::: :::|| : :||:  
 Db 364 FKEISYTT----SFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 419

Qy 412 TLKGAQVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYVL- 470  
 : :|:| | :|: | : :||| || : : : | | | | :|  
 Db 420 ST--GIQNRAGVLFFLTNTNQCFS--VSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLS 476

Qy 471 HVLPFSSVIATVIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFLLTVLLGIVQNPNI 530  
 :|| : : :||: : | : ||| | : | : : : : : | | : :  
 Db 477 DLLPMRMLPSIIFTCIVYFMLGLKPKADAFFVMMFTLM--MVAYSASSMALAIAAGQSV 533

Qy 531 VNSIVALLSIS--GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTC 588  
 | : | :| : :| || : | : : | | ||: :| | ||| ||| |  
 Db 534 VSVATLLMTICFVFMIFSGLLVNLTTIASWLSWLQYFSIPRYGFTALQHNEFLGQNF-C 592

Qy 589 GGSNTSMLNHPMCAITQGVQFIEK 612  
 | | : | | | : : |  
 Db 593 PGLNATGNNPCNYATCTGEEYLVK 616

# RESULT 7

Q96TA8

ID Q96TA8 PRELIMINARY; PRT; 655 AA.  
 AC Q96TA8;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE ATP-binding cassette superfamily G (White) member 2 (Hypothetical protein).  
 GN ABCG2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21201983; PubMed=11306452;  
 RA Komatani H., Kotani H., Hara Y., Nakagawa R., Matsumoto M.,  
 RA Arakawa H., Nishimura S.;  
 RT "Identification of breast cancer resistant protein/mitoxantrone  
 RT resistance/placenta-specific, ATP-binding cassette transporter as a  
 RT transporter of NB-506 and J-107088, topoisomerase I inhibitors with an  
 RT indolocarbazole structure."  
 RL Cancer Res. 61:2827-2832(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreatic carcinoma;

RA Strausberg R.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AB051855; BAB46933.1; -.  
 DR EMBL; BC021281; AAH21281.1; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR006162; Ppantne\_S.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.  
 KW Hypothetical protein; ATP-binding.  
 SQ SEQUENCE 655 AA; 72314 MW; A8AF66B96034C5A8 CRC64;

Query Match 20.5%; Score 691.5; DB 4; Length 655;  
 Best Local Similarity 29.0%; Pred. No. 5.3e-43;  
 Matches 181; Conservative 141; Mismatches 247; Indels 55; Gaps 16;

Qy 25 LEQGSVTGTTEARHS-----LGVLHVSYSVSNRVGPWWNIKSCQQKWDRQILKDV 73  
 : ||: | | | : : | | : | : : : || : :  
 Db 12 VSQGNNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLKSG----FLPCRKPVEKEILSNI 67  
  
 Qy 74 SLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVCGLRRDQFQDCFSYV 133  
 : : : | : ||| : | ||| : || : | : | | : || | : | : ||  
 Db 68 NGIMKPG-LNAILGPTGGGKSSLLDVLAARKDPSG-LSGDVLINGAP-RPANFKCNSGYV 124  
  
 Qy 134 LQSDVFLSSLTVRETLRYTAMLALCRSSADF-YNKKVEAVMTELSLSHVADQMIGSYNFG 192  
 : | || : : |||| | : : | : : | : : | : || | ||| : :  
 Db 125 VQDDVVMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIQELGLDKVADSKVGTQFIR 184  
  
 Qy 193 GISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQ 252  
 | : | ||| : | | : | : || : | ||||| ||| : ||| : : : | : | : |||  
 Db 185 GVSggerKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKRMASKQGRTIIFSIIHQ 244  
  
 Qy 253 PRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCGYPCPEHSNPFDFYMDLTSVDTQ- 311  
 || : | : | : : | | : | | : | || : | | : || | || : : | :  
 Db 245 PRYSIFKLFDSTLLASGRMLMFHGPQAEALGYFESAGYHCEAYNNPADFFLDIINGDSTA 304  
  
 Qy 312 -SREREIETYKRVQMLECAFKESDIYHKI-----LENIERARYLKT 351  
 : || | : | : : : : : : : : | : : :  
 Db 305 VALNRE-EDFKATEIIEPSKQDKPLIEKLAIEYVNSSFYKETKAELHQLSGGEKKKKITV 363  
  
 Qy 352 LPMVPFKTKDPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNN 411  
 : : | : | : : | : || : || : : : : : : : : : : : : : : : :  
 Db 364 FKEISYTT----SFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKND 419  
  
 Qy 412 TLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMRLRAVSDQESQDGLYHKWQMLLAYVL- 470  
 : : : | : : | : : : : : : : : : | | | : : | :  
 Db 420 ST--GIQNRAGVLFFLTNNQCFSS-VSAVELFVVEKKLFIHEYISGYRVSYSYFLGKLLS 476  
  
 Qy 471 HVLFPFSVIATVIFSSVCYWTGLYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNI 530  
 : || : : : || : : | | | : | : : : : : : : : : | | : :  
 Db 477 DLLPMRMLPSIIFTCIVYFMLGLKPKADAFVMMFTLM---MVAYSASSMALAIAAGQSV 533

```

Qy      531 VNSIVALLSIS--GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCC1ILVVNEFYGLNFTC 588
      | :      | ::      : : | | : | :      | | | :      | | | | | |
Db      534 VSVATLLMTICFVFMMIFSGLLVNLTTIASWLSWLQYFSIPRYGFTALQHNEFLGQNF-C 592

Qy      589 GGSNTSMLNHPMCAITQGVQFIEK 612
      | | :      |      | : : |
Db      593 PGLNATGNNPCNYATCTGEEYLVK 616

```

Q8MIB3

```

ID      Q8MIB3      PRELIMINARY;      PRT;      656 AA.
AC      Q8MIB3;
DT      01-OCT-2002 (TrEMBLrel. 22, Created)
DT      01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Brain multidrug resistance protein.
GN      BMDP.
OS      Sus scrofa (Pig).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX      NCBI_TaxID=9823;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=22050127; PubMed=12054514;
RA      Eisenblaetter T., Galla H.J.;
RT      "A new multidrug resistance protein at the blood-brain barrier.";
RL      Biochem. Biophys. Res. Commun. 293:1273-1278(2002).
DR      EMBL; AJ420927; CAD12785.1; -.
DR      PIR; JC7860; JC7860.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0005524; F:ATP binding; IEA.
DR      GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR      GO; GO:0000166; F:nucleotide binding; IEA.
DR      GO; GO:0006810; P:transport; IEA.
DR      InterPro; IPR003593; AAA_ATPase.
DR      InterPro; IPR003439; ABC_transporter.
DR      InterPro; IPR006162; Ppantne_S.
DR      Pfam; PF00005; ABC_tran; 1.
DR      ProDom; PD000006; ABC_transporter; 1.
DR      SMART; SM00382; AAA; 1.
DR      PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
DR      PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
KW      ATP-binding.
SQ      SEQUENCE      656 AA;      72392 MW;      118ADD5B53D9D67F CRC64;

```

Qy	31	TGTEARHSLGVLHVSY-SVSNRVGPWWNIKS---- <td>85</td>	85
		:   :     : : :    :   :   :   :   :   :   :   :	
Db	24	SSNELKTSAGGAVLSFHDICYRV----KVKSGLFCRKTVEKEILTNINGIMKPG-LNAI	78
Qy	86	LGSSSGSGKTTLLDAISGRLLRRTGTLEGEVFNVCGLRRDQFQDCFSYVLQSDVFLSSLTV	145
		:    ::    : :       :  :     :   :     : :	
Db	79	LGPTGGGKSSLLDVLAARKDPHG-LSGDVLINGAP-RPANFKCNSGYVVQDDVVMGTLTV	136



DR EMBL; AY017168; AAG52982.1; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR006162; Ppantne\_S.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.  
 KW ATP-binding.  
 SQ SEQUENCE 655 AA; 72288 MW; B3B5DC02C095C4A8 CRC64;

Query Match 20.3%; Score 683.5; DB 4; Length 655;  
 Best Local Similarity 28.8%; Pred. No. 2.1e-42;  
 Matches 180; Conservative 141; Mismatches 248; Indels 55; Gaps 16;

Qy	25	LEQGSVTGTEARHS-----LGVLHVSYSVSNRVGPWWNIKSCQQKWDRQILKDV	73
		:   :         : :     :     : :    : :	
Db	12	VSQGNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLKSG----FLPCRKPVEKEILSNI	67
Qy	74	SLYIESGQIMCILGSSSGSGKTLLDAISGRLRRTGTLEGEVVFVNGCELRRDQFQDCFSYV	133
		: : :   :     :     :     : :   :       :       :	
Db	68	NGIMKPG-LNAILGPTGGGKSSLLDVLAARKDPSG-LSGDVLINGAP-RPANFKCNSGYV	124
Qy	134	LQSDVFLSSLTVRETLRYTAMLALCRSSADF-YNKKVEAVMTELSLSHVADQMIGSYNFG	192
		:       : :        :    : : :   : :   :           :   :	
Db	125	VQDDVVMGTLTVRENLFSAALRLATTMTNHEKNERINRVIQELGLDKVADSKVGTQFIR	184
Qy	193	GISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQ	252
		:       :     :   :    : :           :     : : :   :   :	
Db	185	GVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKRMSKQGRTIIFSIIHQ	244
Qy	253	PRSELFQHFQDKIAILTYGELVFCGTPPEMLGFFNCGYPCPEHSNPFDFYMDLTSVDTQ-	311
		:   :    : :       :   :   :   :   :   :   :   :   :	
Db	245	PRYSIFKLFDSTLLASGRLMFHGPAQEALGYFESAGYHCEAYNNPADFFLDIINGDSTA	304
Qy	312	-SREREIETYKRVQMLECAFKESDIYHKI-----LENIERARYLKT	351
		:       :   : :   : : :   :   :   :   :   :   :   :   :	
Db	305	VALNRE-EDFKATEIIEPSKQDKPLIEKLAEIYVNSSFYKETKAELHQLSGGEKKKKITV	363
Qy	352	LPMVPFKTKDPPGMFGKLGVLRLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNN	411
		: :   :   : :   :    :       : : :    : : :   : :   :	
Db	364	FKEISYTT----SFCHQLRWVSKRSFKNLLGNPQASIAQIIIVTVVLGLVIGAIYFGLKND	419
Qy	412	TLKGAQVQDRVGLLYQLVGATPYTGMLNAVNLFPMRLRAVSDQESQDGLYHKWQMLLAYVL-	470
		: : :   :   :   : : :      : : :         :	
Db	420	ST--GIQNRAGVLFFLTNQCFS--VSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLS	476
Qy	471	HVLPPFSVIATVIFSSVCYWTGLYPEVARFGYFSAALLAPHLIGEFLLTVLLGIVQNPNI	530
		:    : : :    : :   :    :   : : : : :     : :	
Db	477	DLLPMRMLPSIIFTCIVYFMLGLKAKADAFFVMMFTLM--MVAYSASSMALAIAAGQSV	533

```
Qy      531 VNSIVALLSIS--GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCILVVNEFYGLNFTC 588
      | :      | : : |      : | : | : | : | : | : | : | : | : | : | : | : |
Db      534 VSVATLLMTICFVMMIFSGLLVNLTTIASWLSWLQYFSIPRYGFTALQHNEFLGQNF-C 592

Qy      589 GGSNTSMLNHPMCAITQGVQFIEK 612
      | | : |      |      | : : |
Db      593 PGLNATGNNPCNYATCTGEEYLVK 616
```

Q9R004

```

ID      Q9R004      PRELIMINARY;      PRT;      657 AA.
AC      Q9R004;
DT      01-MAY-2000 (TrEMBLrel. 13, Created)
DT      01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Breast cancer resistance protein 1.
GN      ABCG2 OR BCRP1.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=FVB; TISSUE=Liver;
RX      MEDLINE=99413474; PubMed=10485464;
RA      Allen J.D., Brinkhuis R.F., Wijnholds J., Schinkel A.H.;
RT      "The mouse Bcrp1/Mxr/Abcp gene: amplification and overexpression in
RT      cell lines selected for resistance to topotecan, mitoxantrone, or
RT      doxorubicin.";
RL      Cancer Res. 59:4237-4241(1999).
DR      EMBL; AF140218; AAD54216.1; -.
DR      MGD; MGI:1347061; Abcg2.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0005524; F:ATP binding; IEA.
DR      GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR      GO; GO:0000166; F:nucleotide binding; IEA.
DR      GO; GO:0006810; P:transport; IEA.
DR      InterPro; IPR003593; AAA_ATPase.
DR      InterPro; IPR003439; ABC_transporter.
DR      InterPro; IPR006162; Ppantne_S.
DR      Pfam; PF00005; ABC_tran; 1.
DR      ProDom; PD000006; ABC_transporter; 1.
DR      SMART; SM00382; AAA; 1.
DR      PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
DR      PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
KW      ATP-binding.
SQ      SEQUENCE      657 AA;      73021 MW;      207B70BC272CC0D5 CRC64;

```

Query Match 20.1%; Score 677; DB 11; Length 657;  
Best Local Similarity 29.7%; Pred. No. 6.4e-42;  
Matches 189; Conservative 135; Mismatches 236; Indels 76; Gaps 21;

QY            15 PHINRGLSSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS---CQQKWDRQILK 71  
             | :| :: :| :| |                     | :::| |                     :||        :: :::||  
Db            21 PRMNSRAVRTLAEGDV-----LSFHHITYRV-----KVKSGFLVRKTVEKEILS 64



Qy 72 DVSLYIESGQIMCILGSSSGKTTLLDAISGRLLRRTGTLEGEVFNVCCELRRDQFQDCFS 131  
 |:: :: | : ||| :| ||::||| :: | | | :| :| : | : |  
 Db 65 DINGIMKPG-LNAILGPTGGGKSSLLDVLAARKDPKG-LSGDVLINGAP-QPAHFKCCSG 121

Qy 132 YVLQSDVFLSSLTVRETLRYTAMLALCRSSADF-YNKKVEAVMTLSLSHVADQMIGSYN 190  
 ||:| || : :||| ||::| | | : : |::: :: || | ||| :|:  
 Db 122 YVQDDVVMGTLTVRENLFSAALRLPTTMKNHEKNERINTI IKELGLEKVADSKVGTQF 181

Qy 191 FGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAEELARRDRIVIVTI 250  
 ||| |||:| || :|: || :: ||||| ||| ::||| :::: | :| :|  
 Db 182 IRGISGGERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVLLLLKRMSKQGRTIIFSI 241

Qy 251 HQPRSELFQHFDKIAILTYGELVFCGTPEEMLGFFNCCGYPCPEHSNPFDFYMDLTSVDT 310  
 |||| :|: || : :| |::|| | : : | : | : || | :|| ||::|: : |:  
 Db 242 HQPRYSIFKLFDLTLASGKLVFHGPAQKALEYFASAGYHCEPYNNPADFFLDVINGDS 301

Qy 311 QS-----REREIETYKR-----VQMLECAFKESDIYHKILENIERARYLKTLPMV 355  
 : |:: | | : : | : | | | | | | | |  
 Db 302 SAVMLNREEQDNEANKTEEPSKGEKPVIENTLSEFYINSAIYG-----ETKAELDQLPGA 355

Qy 356 PFKT-----KDP---PGMFGKLGVLRLRVTRNLMRNKQAVIMRLVQNLIMGLFL--IFYL 405  
 | :| :| :||| | | : :| :|:| : |::  
 Db 356 QEKKGTSAFKEPVYVTSFCHQLRWIARRSFKNLLGNPQASVAQLIVTVILGLIIGAIYFD 415

Qy 406 LRVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQML 465  
 |: :|:| |::| : : ::|| | | : : : | | |  
 Db 416 LKYD----AAGMQNRAGVLFLLTTNQCFSS-VSAVELFVVEKKLFIHEYISGYRVSSYF 470

Qy 466 LAYVL-HVLPFSVIATVIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFLLVLLGI 524  
 |: :|| : :|||: : |: ||| | | | : : : : | |  
 Db 471 FGKVMSDLPLMRELPVIFTCILYFMLGLKKTVDAFFIMFTLI---MVAYTASSMALAI 527

Qy 525 VQNPNIIVNSIVALLSIS--GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCILVNEFY 582  
 ::|: |::|: ::|| : |:: : | | ||: :| | |||  
 Db 528 ATGQSVSVATLLMTIAFVFMMLFSGLLVNLRITIGPWLSQLQYFSIPRYGFTALQYNEFL 587

Qy 583 GLNFTCGGSN-----TSMLNHPMCA-----ITQGVQ 608  
 | | | | | | : : : | | | : :  
 Db 588 QGEF-CPGFNVTDNSTCVNSYAICTGNEYLINQGIE 622

# RESULT 11

Q7TMS5

ID Q7TMS5 PRELIMINARY; PRT; 657 AA.  
 AC Q7TMS5;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE ATP-binding cassette, sub-family G, member 2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6Ncr; TISSUE=Hematopoietic Stem Cell;  
 RX MEDLINE=22388257; PubMed=12477932;

Qy	15	PHINRGLSSLSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS---	CQQKWDQRQILK	71
		: :   :     : :   :   : :		
Db	21	PRTNSRAVRTLAEGDV-----LSFHHITYRV-----	KVKSGLVLRKTVEKEILS	64
Qy	72	DVSLYIESGQIMCILGSSSGSKTTLLDAISGRRLRRTGTLEGEVFN	GCELRDQFQDCFS	131
		: : :   :     :      :     : :       :   :   :		
Db	65	DINGIMKPG-LNAILGPTGGGKSSLLDVLAARKDPKG-LSGDVLINGAP-Q	PAHFKCCSG	121
Qy	132	YVLQSDVFLSSLTVRETLRYTAMLALCRSSADF-YNKKVEAVMT	ELSLSHVADQMIGSYN	190
		:     : :       :    :     : :    : :     :   :		
Db	122	YVVQDDVVMGTLTVRENLFSAALRLPTTMKNHEKNERINTI	IKELGLEKVDASKVGTQF	181
Qy	191	FGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIV	LLLAELARRDRIVVTI	250
		:      :   :    :           : :     : :   :   :		
Db	182	IRGISGGERKRTSIGMELITDPSILFLDEPTTGLDSSSTANAV	LLLLKRMSKQGRTIIFSI	241
Qy	251	HQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFN	CGYPCPEHSNPFDFYMDLTSVDT	310
		: :     : :           : :   :     :       : :   :   :		
Db	242	HQPRYSIFKLFDSLTLLASGKLVFHGPAQKALEYFASAGYHCEP	YNPNPADFFLDVINGDS	301
Qy	311	QS-----REREIETYKR-----	VQMLECAFKESDIYHKILENIERARYLKTLP	355
		:   : :     : :   :		
Db	302	SAVMLNREEQDNEANKTEEPSKGEKPVIENTLSEFYINS	AIYG-----ETKAELDQLPGA	355

Qy 356 PFKT-----KDP---PGMFGKLGVLRLRRVTRNLMRNKQAVIMRLVQNLIMGLFL--IFYL 405  
 | | | | | : | : | | : | | : : | : | | : | :  
 Db 356 QEKKGTSFAKPEPVYVTSFCHQLRWIARRSFKNLLGNPQASVAQLIVTVILGLIIGAIYFD 415  
 Qy 406 LRVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQML 465  
 | : | : | | | : | : | : | | : : : | | |  
 Db 416 LKYD-----AAGMQNRAGVLFLLTTNQCFSS-VSAVELFVVEKKLFIHEYISGYRVSSYF 470  
 Qy 466 LAYVL-HVLPFSVIATVIFSSVCYWTGLYPEVARFGYFSAALLAPHLIGEFITLVLLGI 524  
 | : : | | : : | | | | : | | | | : : : : | |  
 Db 471 FGKMSDLLPMRFLPSVIFTCLVLYFMLGLKKTVDAFFIMMFTLI---MVAYTASSMALAI 527  
 Qy 525 VQNPNIIVNSIVALLSIS--GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFY 582  
 : : | : | : | : : : | : : : | | | : | | |  
 Db 528 ATGQSVSVATLLMTIAFVFMMLFSGLLVNLRTIGPWLSWLQYFSIPRYGFTALQYNEFL 587  
 Qy 583 GLNFTCGGSN-----TSMNLNHPMCA-----ITQGVQ 608  
 | | | | | : : : | | | : | | :  
 Db 588 GQEF-CPGFNVTDNSTCVNSYAICTGNEYLINQGIE 622

RESULT 12

Q80W57

ID Q80W57 PRELIMINARY; PRT; 657 AA.  
 AC Q80W57;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE ABC transporter ABCG2.  
 GN ABCG2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=wistar;  
 RA Hori S., Ohtsuki S., Terasaki T.;  
 RT "Expression and regulation of ABCG2 at the rat blood-brain barrier."  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AB105817; BAC76396.1; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR006162; Ppantne\_S.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.  
 SQ SEQUENCE 657 AA; 72960 MW; C975C61A08489027 CRC64;

Query Match

19.6%; Score 659; DB 11; Length 657;

Best Local Similarity 29.3%; Pred. No. 1.4e-40;  
Matches 192; Conservative 132; Mismatches 248; Indels 84; Gaps 21;

```

Qy      4 LPFLSPEGARGPHINRGSLSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKSCQQ 63
      || :| |||      :| :| |      | | :|| | : | : :: :
Db     20 LPGMSSRGAR-----TLAEGDV-----LSFHHITYRVKVKSG--FLVRKTAE 59

Qy     64 KWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVCCELRR 123
      |  :|| |:: :: | : ||| :| ||::||| :: |  | | | :| | :
Db     60 K---EILSDINGIMKPG-LNAILGPTGGGKSSLLDVLAAKDPKG-LSGDVLINGAP-QP 113

Qy    124 DQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRS-SADFYNKKVEAVMTELSLHVA 182
      | :  ||:| || : :||| | |::| | | :  |::: :: || | ||
Db    114 ANFKCSSGYVQDDVVMGTLTVRENLQFSAAALRLPKAMKTHEKNERINTI IKELGLDKVA 173

Qy    183 DQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARR 242
      | :| :  ||| |||:| || :| :| :| :| | | | | | | | | | | | | | | | |
Db    174 DSKVGTQFTRGISGGERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVLLLLKRMSKQ 233

Qy    243 DRIVIVTIHQPRSELFQHFDKIAILTYGELVFCGTPEEMLGFFNCCGYPCPEHSNPFDFY 302
      | :| :||| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db    234 GRTIIFSIIHQPRYSIFKLFDLSLTLASGKLMFHGPAQKALEYFASAGYHCEPYNNPADFF 293

Qy    303 MDLTSVDTQS-----REREIETYKRVQMLECAFKESDIYHKILENI-----ERA 346
      :| : : | : :  :| : | | :  |  | :|| :  |
Db    294 LDVINGDSSAVMLNRGEQDHEANKTEE-----PSKREKPIIENLAEFYINSTIYGETK 346

Qy    347 RYLKTLPMVPFKT-----KDP----PGMFGKLGVLRLRRVTRNLMRNKQAVIMRLVQNLMG 398
      | ||: |  :||  :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db    347 AELDQLPVAQKKKGSSAFREPVVYVTSFCHQLRWIARRSFKNLLGNPQASVAQLIVTVILG 406

Qy    399 LFLIFYLLRVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGL 458
      | :  :|| :  :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db    407 LIIGALYFGLKNDPT--GMQNRAGVFFFLTTNQCFTS-VSAVELFVVEKKLFIHEYISGY 463

Qy    459 YHKWQMLLA-YVLHVLFPFSVIATVIFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGEF 517
      |  :|| :  :||: :| :| | | | | | | | | | | | | | | | : :
Db    464 YRVSSYFFGKLVSDLLPMRFLPSVIYTCILYFMLGLKRTVEAFFIMFTLI---MVAYTA 520

Qy    518 TLVLLGIVQNPNIIVNSIVALLSIS--GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEI 575
      : : | |  :| :  :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db    521 SSMALAIAGQSVSVATLLMTISFVFMMLFSGLLVNLRTIGPWLSWLQYFSIPRYGFTA 580

Qy    576 LVVNEFYGLNFTCGGSNTSMLNHPMCAITQGVQFIEKTCPGATSRFTAN-FLILYG 630
      | ||| | | | | | :|  | | : : | | :| | |
Db    581 LQHNEFLGQEF-CPGLNVTM-----NSTCVNSYTICTGNDYLINQG 620

```

# RESULT 13

Q80XF3

ID Q80XF3 PRELIMINARY; PRT; 657 AA.

AC Q80XF3;

DT 01-JUN-2003 (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE ATP-binding cassette transporter ABCG2.

GN ABCG2.



```

      | ||:      || |:|      :| : || :||: | || : |: :|
Db      347 AELDQLPVAQKKKGSSPF--KEPVYVTSFCHQLRWIARRSFKNLLGNPQASVAQLIVTVI 404

Qy      397 MGLFLIFYLLRVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMRLRAVSDQESQD 456
      :|| :      :||:      :||: | : |      :| :|| || : : : |
Db      405 LGLIIGALYFGLKNDPT--GMQNRAGVFLEFLTNNQCFTS-VSAVELFVVEKKLFIHEYIS 461

Qy      457 GLYHKWQMLLA-YVLHVLPPFSVIATVIFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGE 515
      | |      | :|| : :||: : : | : || | | | : : :
Db      462 GYYRVSSYFFGKLVSDLLPMRFLPSVIYTCILYFMLGLKRLVEAFFIMRFTLI---MVAY 518

Qy      516 FLTIVLLGIVQNPNIIVNSIVALLSIS--GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCC 573
      : : | | : :|| : :|| : :|| : :|| : :|| : :|| : :||
Db      519 TASSMALAIAAGQSVVSVATLLMTISFVFMMLFSGLLVNLRITGPWLSWLQYFSIPRYGF 578

Qy      574 EILVVNEFYGLNFTCGGSNTSMLNHPMCAITQGVQFIEKTCPGATSRFTAN-FLILYG 630
      | ||| | | | | :|      || : : | | :|| |
Db      579 TALQHNEFLGQEF-CPGLNVTM-----NSTCVNSYTICTGNDYLINQG 620

```

RESULT 14

Q80ST1

```

ID   Q80ST1          PRELIMINARY;          PRT;    657 AA.
AC   Q80ST1;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   ATP-binding cassette protein G2 transcript variant B (ATP-binding
DE   cassette protein G2 transcript variant C) (ATP-binding cassette
DE   protein G2 transcript variant A).
GN   ABCG2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RA   Yabuuchi H., Ishikawa T.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY089996; AAM09106.1; -.
DR   EMBL; AY089997; AAM09107.1; -.
DR   EMBL; AY089998; AAM09108.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR   GO; GO:0000166; F:nucleotide binding; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR006162; Ppantne_S.
DR   Pfam; PF00005; ABC_tran; 1.
DR   ProDom; PD000006; ABC_transporter; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
KW   ATP-binding.

```

SQ SEQUENCE 657 AA; 72960 MW; E194871E1C1AC201 CRC64;

Query Match 19.5%; Score 657; DB 11; Length 657;  
Best Local Similarity 29.3%; Pred. No. 2e-40;  
Matches 192; Conservative 132; Mismatches 248; Indels 84; Gaps 21;

```
Qy      4 LPFLSPEGARGPHINRGSLSSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKSCQQ 63
      || :| |||          :| :| |          |   |::| | : | : :: :
Db      20 LPGMSSRGAR-----TLAEGDV-----LSFHHITYRVKVKSG--FLVRKTAE 59

Qy      64 KWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVCCELRR 123
      |  :|| |:: :| : ||| :| ||::||| :: |   | | |::| :|| :
Db      60 K---EILSDINGIMKPG-LNAILGPTGGGKSSLLDVLAARKDPRG-LSGDVLINGAP-QP 113

Qy     124 DQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRS-SADFYNNKKVEAVMTLSLHVA 182
      | :   ||:| || : :||| ||::| | | ::   |:: : : || | ||
Db     114 ANFKCSSGYVQDDVVMGTLTVRENLFSAALRLPKAMKTHEKNERINTIIKELGLDKVA 173

Qy     183 DQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARR 242
      | :| : || | ||:| || :| : || : ||| ||| ||| :||| ::::
Db     174 DSKVGTQFTRGISGGERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVLLLLKRMSKQ 233

Qy     243 DRIVIVTIHQPRSELFQHFDKIAILTYGELVFCGTPEEMLGFFNNGGYPCPEHSNPFDFY 302
      | :| :||| | :| : || :| |::| | : :| :| : || | :|| ||:
Db     234 GRTIIFSIHQPRYSIFKLFDSLTLASGKLMFHGPAQKALEYFASAGYHCEPYNNPADFF 293

Qy     303 MDLTSVDTQS-----REREIETYKRVQMLECAFKESDIYHKILENI-----ERA 346
      :| : : | : :   |:: | | :   |   |::| :   |
Db     294 LDVINGDSSAVMLNRGEQDHEANKTEE-----PSKREKPIIENLAEFYINSTIYGETK 346

Qy     347 RYLKTLPMVPFKT-----KDP---PGMFGKLGVLRLRRVTRNLMRNKQAVIMRLVQNLIMG 398
      | ||: |   :||   :| : || :||: | || : :| : :||
Db     347 AELDQLPVAQKKKGSSAFREPVVYTSFCHQLRWIARRSFKNLLGNPQASVAQLIVTVILG 406

Qy     399 LFLIFYLLRVQNNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGL 458
      | :   :| :   :| : | : |   :| :|| || : : : | |
Db     407 LIIGALYFGLKNDPT--GMQNRAGVFFFLTNNQCFTS-VSAVELFVVEKKLFIHEYISGY 463

Qy     459 YHKWQMLLA-YVLHVLPPFSVIATVIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEF 517
      |   | :|| : :||: : | : || | | |   | :   ::
Db     464 YRVSSYFFGKLVSDLLPMRFLPSVIYTCLLYFMLGLKRTVEAFFIMMFTLI---MVAYTA 520

Qy     518 TLVLLGIVQNPNIIVNSIVALLSIS--GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEI 575
      : : | |   :| :   |::||   :: || : |:: : | | ||: :|
Db     521 SSMALAIAGQSVSVSATLLMTISFVEMMLFSGLLVNLRTIGPWLSWLQYFSIPRYGFTA 580

Qy     576 LVVNEFYGLNFTCGGSNTSMLNHPMCAITQGVQFIEKTCPGATSRFTAN-FLILYG 630
      | ||| | | | | :|   || : : | | :|| |
Db     581 LQHNEFLGQEF-CPGLNVTM-----NSTCVNSYTICTGNDYLINQG 620
```

RESULT 15

Q8T691

ID Q8T691 PRELIMINARY; PRT; 801 AA.

AC Q8T691;

DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE ABC transporter AbcG1.  
 GN ABCG1.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ax4;  
 RA Anjard C., Loomis W.F.;  
 RT "Evolution of the ABC transporters of Dictyostelium.";  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.  
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
 DR EMBL; AF482380; AAL91485.1; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 KW ATP-binding; Transport.  
 SQ SEQUENCE 801 AA; 90052 MW; CCC4F0036CB195A3 CRC64;

Query Match 19.0%; Score 640; DB 5; Length 801;  
 Best Local Similarity 27.5%; Pred. No. 4.7e-39;  
 Matches 188; Conservative 126; Mismatches 249; Indels 120; Gaps 19;

Qy	59	KSCQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLLRRTGTLEGEVFN	118
		:  :	
Db	128	KGKKKKISKQILTNINGHIESGTIFAIMGPSGAGKTTLLDILAHRLNINGS--GTMVLYNG	185
Qy	119	CELRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNK--KVEAVMTEL	176
		: :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	
Db	186	NKSDFNIFKKLCGYVTQSDSLMPSLTVRETLNFYAQLKMPR-DVPLKEKLQRVQDIIDEM	244
Qy	177	SLSHVADQMIGSYN--FGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVL	234
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	
Db	245	GLNRCADTLVGTADNKIRGISGGERRRVTISIELLTGPSVILLDEPTSGLDASTSFYVMS	304
Qy	235	LLAELARRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCGYPCPE	294
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	
Db	305	ALKKLAKSGRTIICTIHQPRSNIDMFDNLLLLGDGNTIYYGKANKALEYFNANGYHCSE	364
Qy	295	HSNPFDFYMDL--TSVDTQS-----REREI-----	317
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	
Db	365	KTNPADFFLDLINTQVEDQADSDDDDDYNDEEEEIGGGGGSGGGAGGIEDIGISISPTMN	424
Qy	318	-----ETYKRVQMLECAFKESDIYHKILENIERA-RY	348
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :	
Db	425	GSAVDNIKNNELKQQQQQQQQQQQSTDGRRARRRIKKL--TKEEMVILKKEYPNSEQGLRV	482



Qy 349 LKTLPMV-----PFKTKDPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLF 400  
 :|| : : : | : :|| | | |: | : |:| : ||  
 Db 483 NETLDNISKENRTDFKYEKTRGPNFLTQFSLLLGREVTNAKRHPMAFKVNLIQAIQGLL 542

Qy 401 --LIFYLLRVQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGL 458  
 :::| | : : :|| | |:: : : : : :|| : : : : |:  
 Db 543 CGIVYYQLGLG---QSSVQSRTGVVAFIIMGVSEFPAVMSTIHVFPDVITIFLKDRASGV 598

Qy 459 YHKWQMLLAYVLHVL PFSVIATVIFSSVCYWTLG-----LYPEVARFGYFSAALLAPHLI 513  
 | || :|: : : : || | | : : ||  
 Db 599 YDTLPFFLAKSFMDACIAVLLPMVTATIVYWMTNQRVDPFYSAAPFFRFVLMLVLASQ-- 656

Qy 514 GEFLTLVLLGIVQN---PNI-VNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQ 569  
 | : ||:: : ||: | : || | : : ||| |: : :| | | :|  
 Db 657 ----TCLSLGVLISSSVPNVQVGTAVAPLIVILFFLFSGFFINLNDVPGWLWVFPYISFF 712

Qy 570 KYCCEILVVNEFYGLNFTC-----GGSNTSMLNHPMCAITQGVQFIEKTCPGATSRFTA 623  
 :| | |:| | :||| || :| : | || |  
 Db 713 RYMIEAAVINAFKDVHFTCTDSQKIGG-----VCPVQYGNNVIE-NMGYDIDHFWR 762

Qy 624 NFLILYGFIPALVILGIVIFKVR 646  
 | || :| :| : :|:  
 Db 763 NVWILVLYIIGFRVLTFLVLKLK 785

Search completed: February 27, 2004, 07:15:25  
 Job time : 40.1949 secs

OM protein - protein search, using sw model

Run on: February 27, 2004, 06:40:43 ; Search time 10.0952 Seconds  
(without alignments)  
3362.970 Million cell updates/sec

Title: US-09-989-981A-2  
Perfect score: 3369  
Sequence: 1 MGELPFLSPEGARGPHINRG.....PALVILGIVIFKVRDYILSR 652

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query	Match Length			
1	3363	99.8	652	1	ABG5_MOUSE	Q99pe8 mus musculu
2	3144	93.3	652	1	ABG5_RAT	Q99pe7 rattus norv
3	2744.5	81.5	651	1	ABG5_HUMAN	Q9h222 homo sapien
4	720.5	21.4	694	1	ABG8_RAT	P58428 rattus norv
5	701.5	20.8	673	1	ABG8_MOUSE	Q9dbm0 mus musculu
6	690.5	20.5	655	1	ABG2_HUMAN	Q9unq0 homo sapien
7	688.5	20.4	673	1	ABG8_HUMAN	Q9h221 homo sapien
8	628	18.6	1294	1	YOH5_YEAST	Q08234 saccharomyc
9	602.5	17.9	1049	1	ADP1_YEAST	P25371 saccharomyc
10	589	17.5	687	1	WHIT_DROME	P10090 drosophila
11	586	17.4	678	1	ABG1_HUMAN	P45844 homo sapien
12	584.5	17.3	598	1	YPC3_CAEEL	Q11180 caenorhabdi
13	583.5	17.3	677	1	WHIT_LUCCU	Q05360 lucilia cup
14	570.5	16.9	695	1	WHIT_ANOGA	Q27256 anopheles g
15	569.5	16.9	646	1	ABG4_HUMAN	Q9h172 homo sapien
16	568.5	16.9	666	1	ABG1_MOUSE	Q64343 mus musculu
17	560	16.6	666	1	SCRT_DROME	P45843 drosophila

18	557	16.5	679	1	WHIT_CERCA	Q17320	ceratitis c
19	554	16.4	610	1	YQ5C_CAEEL	Q09466	caenorhabdi
20	530.5	15.7	709	1	WHIT_ANOAL	Q16928	anopheles a
21	465.5	13.8	675	1	BROW_DROME	P12428	drosophila
22	457	13.6	1333	1	YN99_YEAST	P53756	saccharomyc
23	454.5	13.5	668	1	BROW_DROVI	Q24739	drosophila
24	446	13.2	1499	1	CDR2_CANAL	P78595	candida alb
25	426.5	12.7	1501	1	CDR3_CANAL	O42690	candida alb
26	424.5	12.6	650	1	ABG3_MOUSE	Q99p81	mus musculu
27	416.5	12.4	1529	1	PDRF_YEAST	Q04182	saccharomyc
28	406.5	12.1	1564	1	PDRF_YEAST	P51533	saccharomyc
29	406	12.1	1501	1	SNQ2_YEAST	P32568	saccharomyc
30	405.5	12.0	1511	1	PDR5_YEAST	P33302	saccharomyc
31	405.5	12.0	1530	1	BFR1_SCHPO	P41820	schizosacch
32	396.5	11.8	1410	1	PDRB_YEAST	P40550	saccharomyc
33	394	11.7	1501	1	CDR1_CANAL	P43071	candida alb
34	393	11.7	1511	1	PDRF_YEAST	Q02785	saccharomyc
35	385.5	11.4	1490	1	CDR4_CANAL	O74676	candida alb
36	259.5	7.7	1704	1	ABC3_HUMAN	Q99758	homo sapien
37	258	7.7	241	1	YHBG_HAEIN	P45073	haemophilus
38	256.5	7.6	365	1	CYSA_ECO57	Q8xbj8	escherichia
39	256.5	7.6	365	1	CYSA_ECOL6	Q8ffb3	escherichia
40	256.5	7.6	365	1	CYSA_ECOLI	P16676	escherichia
41	250	7.4	573	1	CYDC_ECOLI	P23886	escherichia
42	248.5	7.4	364	1	CYSA_SALTI	Q8z4v6	salmonella
43	248.5	7.4	365	1	CYSA_SALTY	P40860	salmonella
44	248	7.4	308	1	NOSF_PSEST	P19844	pseudomonas
45	247.5	7.3	231	1	YTRE_BACSU	O34392	bacillus su

# ALIGNMENTS

## RESULT 1

### ABG5\_MOUSE

ID ABG5\_MOUSE STANDARD; PRT; 652 AA.  
AC Q99PE8;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).  
GN ABCG5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Liver;  
RX MEDLINE=20578753; PubMed=11138003;  
RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,  
RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,  
RA Dean M., Patel S.B.;  
RT "Identification of a gene, ABCG5, important in the regulation of  
RT dietary cholesterol absorption."  
RL Nat. Genet. 27:79-83(2001).  
RN [2]

RP TISSUE SPECIFICITY, AND INDUCTION.  
RX MEDLINE=20553648; PubMed=11099417;  
RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,  
RA Kwiterovich P., Shan B., Barnes R., Hobbs H.H.;  
RT "Accumulation of dietary cholesterol in sitosterolemia caused by  
RT mutations in adjacent ABC transporters.";  
RL Science 290:1771-1775(2000).  
CC -!- FUNCTION: Transporter that appears to play an indispensable role  
CC in the selective transport of the dietary cholesterol in and out  
CC of the enterocytes and in the selective sterol excretion by the  
CC liver into bile.  
CC -!- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to  
CC ABCG8 along a pathway regulating dietary-sterol absorption and  
CC excretion (By similarity).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
CC -!- TISSUE SPECIFICITY: Expressed in the intestine and, at lower  
CC level, in the liver.  
CC -!- INDUCTION: Upregulated by cholesterol feeding. Possibly mediated  
CC by the liver X receptor/retinoic X receptor (LXR/RXR) pathway.  
CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
CC subfamily.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF312713; AAG53097.1; -.  
DR MGD; MGI:1351659; Abcg5.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003439; ABC\_transporter.  
DR Pfam; PF00005; ABC\_tran; 1.  
DR ProDom; PD000006; ABC\_transporter; 1.  
DR SMART; SM00382; AAA; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
KW ATP-binding; Glycoprotein; Transmembrane; Transport.  
FT DOMAIN 1 385 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 386 406 1 (POTENTIAL).  
FT DOMAIN 407 422 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 423 443 2 (POTENTIAL).  
FT DOMAIN 444 463 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 464 484 3 (POTENTIAL).  
FT DOMAIN 485 504 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 505 525 4 (POTENTIAL).  
FT DOMAIN 526 529 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 530 550 5 (POTENTIAL).  
FT DOMAIN 551 622 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 623 643 6 (POTENTIAL).  
FT DOMAIN 644 652 CYTOPLASMIC (POTENTIAL).  
FT NP\_BIND 87 94 ATP (POTENTIAL).  
FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 585 585 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 652 AA; 73244 MW; 80CE37ADCC19771E CRC64;

Query Match 99.8%; Score 3363; DB 1; Length 652;  
Best Local Similarity 99.8%; Pred. No. 7.8e-235;  
Matches 651; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 MGELPFLSPEGARGPHINRGSLSSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS 60
        |||
Db      1 MGELPFLSPEGARGPHINRGSLSSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS 60

Qy     61 CQQKWDRQILKDVSPLYIESGQIMCILGSSSGSKTTLLDAISGRLRRTGTLEGEVFNVC 120
        |||
Db     61 CQQKWDRQILKDVSPLYIESGQIMCILGSSSGSKTTLLDAISGRLRRTGTLEGEVFNVC 120

Qy    121 LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTLSLSH 180
        |||
Db    121 LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTLSLSH 180

Qy    181 VADQMIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240
        |||
Db    181 VADQMIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240

Qy    241 RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCGYPCPEHSNPF 300
        |||
Db    241 RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCGYPCPEHSNPF 300

Qy    301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPFKTK 360
        |||
Db    301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPFKTK 360

Qy    361 DPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNTLKGAVQDR 420
        |||
Db    361 DPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNTLKGAVQDR 420

Qy    421 VGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLPPFSVIAT 480
        |||
Db    421 VGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLPPFSVIAT 480

Qy    481 VIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPINVSIVALLSI 540
        |||
Db    481 VIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPINVSIVALLSI 540

Qy    541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPM 600
        |||
Db    541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPM 600

Qy    601 CAITQGVQFIEKTCPGATSRFTANFLILYGFIPALVILGIVIFKVRDYLI 652
        |||
Db    601 CAITQGVQFIEKTCPGATSRFTANFLILYGFIPALVILGIVIFKVRDYLI 652
```

RESULT 2

ABG5\_RAT

ID ABG5\_RAT STANDARD; PRT; 652 AA.

AC Q99PE7; Q8CIQ4;

DT 28-FEB-2003 (Rel. 41, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).  
 GN ABCG5.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Small intestine;  
 RX MEDLINE=20578753; PubMed=11138003;  
 RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,  
 RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,  
 RA Dean M., Patel S.B.;  
 RT "Identification of a gene, ABCG5, important in the regulation of  
 RT dietary cholesterol absorption.";   
 RL Nat. Genet. 27:79-83(2001).  
 RN [2]  
 RP REVISION TO 2.  
 RA Lu K., Lee M.-H., Patel S.B.;  
 RL Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND VARIANT CYS-583.  
 RC STRAIN=GH, SHR, SHRSP, Sprague-Dawley, Wistar, Wistar Kyoto, and WKA;  
 RX PubMed=12783625;  
 RA Yu H., Pandit B., Klett E., Lee M.H., Lu K., Helou K., Ikeda I.,  
 RA Egashira N., Sato M., Klein R., Batta A., Salen G., Patel S.B.;  
 RT "The rat STSL locus: characterization, chromosomal assignment, and  
 RT genetic variations in sitosterolemic hypertensive rats.";   
 RL BMC Cardiovasc. Disord. 3:4-4(2003).  
 CC -!- FUNCTION: Transporter that appears to play an indispensable role  
 CC in the selective transport of the dietary cholesterol in and out  
 CC of the enterocytes and in the selective sterol excretion by the  
 CC liver into bile.  
 CC -!- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to  
 CC ABCG8 along a pathway regulating dietary-sterol absorption and  
 CC excretion (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -!- TISSUE SPECIFICITY: Expressed only in liver and intestine.  
 CC -!- POLYMORPHISM: The polymorphism at position 583 is found in strains  
 CC SHR, SHRSP and Wistar Kyoto which are both hypertensive and  
 CC sitosterolemic. Strains which are hypertensive but not  
 CC sitosterolemic do not contain a polymorphism at this position.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
 CC subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF312714; AAG53098.3; -.   
 DR EMBL; AY145899; AAN64275.1; -.   
 DR InterPro; IPR003593; AAA\_ATPase.



```

Db      361 NPPGMFCKLGVLRRVTRNLMRNKQVVIMRLVQNLIMGLFLIFYLLRVQNNMLKGA VQDR 420
Qy      421 VGLLYQLVGATPYTGMLNAVNLFPM LRAVSDQESQDGLYHKWQMLLAYVLHVPF SVIAT 480
        |||||:|
Db      421 VGLLYQLVGATPYTGMLNAVNLFPM LRAVSDQESQDGLYQKWQMLLAYVLHALPFS IVAT 480
Qy      481 VIFSSVCYWTLGLYPEVARFGYFSAALLAPHLIGEF LTLVLLGIVQNPNI VNSIVALLSI 540
        |||||:|
Db      481 VIFSSVCYWTLGLYPEVARFGYFSAALLAPHLIGEF LTLVLLGMVQNPNI VNSIVALLSI 540
Qy      541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILV VNEFYGLNFTCGGSNTSMLNHPM 600
        |||||:|
Db      541 SGLLIGSGFIRNIEEMPIPLKILGYFTFQKYCCEILV VNEFYGLNFTCGGSNTSV PNNPM 600
Qy      601 CAITQGVQFIEKTCPGATSRFTANFLILYGFIPALV ILGIVIFKVRDY LISR 652
        |:|:|:|
Db      601 CSMTQGIQFIEKTCPGATSRFTTNFLILYSFIPTLV ILGMVVFKVRDY LISR 652

```

# RESULT 3

## ABG5\_HUMAN

```

ID      ABG5_HUMAN      STANDARD;      PRT;      651 AA.
AC      Q9H222;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      ATP-binding cassette, sub-family G, member 5 (Sterolin-1).
GN      ABCG5.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A., AND VARIANT GLU-604.
RC      TISSUE=Liver;
RX      MEDLINE=20553648; PubMed=11099417;
RA      Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
RA      Kwiterovich P., Shan B., Barnes R., Hobbs H.H.;
RT      "Accumulation of dietary cholesterol in sitosterolemia caused by
RT      mutations in adjacent ABC transporters.";
RL      Science 290:1771-1775(2000).
RN      [2]
RP      SEQUENCE FROM N.A., VARIANTS SITOSTEROLEMIA HIS-389; HIS-419 AND
RP      PRO-419, AND VARIANT GLU-604.
RC      TISSUE=Liver;
RX      MEDLINE=20578753; PubMed=11138003;
RA      Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,
RA      Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,
RA      Dean M., Patel S.B.;
RT      "Identification of a gene, ABCG5, important in the regulation of
RT      dietary cholesterol absorption.";
RL      Nat. Genet. 27:79-83(2001).
RN      [3]
RP      REVIEW.
RX      MEDLINE=21474438; PubMed=11590207;
RA      Schmitz G., Langmann T., Heimerl S.;
RT      "Role of ABCG1 and other ABCG family members in lipid metabolism.";

```



RL J. Lipid Res. 42:1513-1520(2001).  
 RN [4]  
 RP VARIANTS SITOSTEROLEMIA GLN-146; HIS-389; PRO-419; HIS-419 AND  
 RP SER-550, AND VARIANT GLU-604.  
 RX MEDLINE=21344600; PubMed=11452359;  
 RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,  
 RA Ose L., Stalenhoef A.F.H., Mietinnen T., Bjorkhem I., Bruckert E.,  
 RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,  
 RA Patel S.B.;  
 RT "Two genes that map to the STSL locus cause sitosterolemia: genomic  
 RT structure and spectrum of mutations involving sterolin-1 and  
 RT sterolin-2, encoded by ABCG5 and ABCG8, respectively.";  
 RL Am. J. Hum. Genet. 69:278-290(2001).  
 CC -!- FUNCTION: Transporter that appears to play an indispensable role  
 CC in the selective transport of the dietary cholesterol in and out  
 CC of the enterocytes and in the selective sterol excretion by the  
 CC liver into bile.  
 CC -!- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to  
 CC ABCG8 along a pathway regulating dietary-sterol absorption and  
 CC excretion.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -!- TISSUE SPECIFICITY: Strongly expressed in the liver, lower levels  
 CC in the small intestine and colon.  
 CC -!- DISEASE: Defects in ABCG5 are a cause of sitosterolemia  
 CC [MIM:210250]; also known as phytosterolemia or shellfish  
 CC sterolemia. It is a rare autosomal recessive disorder  
 CC characterized by increased intestinal absorption of all sterols  
 CC including cholesterol, plant and shellfish sterols, and decreased  
 CC biliary excretion of dietary sterols into bile. Sitosterolemia  
 CC patients have hypercholesterolemia, very high levels of plant  
 CC sterols in the plasma, and frequently develop tendon and tuberous  
 CC xanthomas, accelerated atherosclerosis and premature coronary  
 CC artery disease.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
 CC subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF320293; AAG40003.1; -.  
 DR EMBL; AF312715; AAG53099.1; -.  
 DR Genew; HGNC:13886; ABCG5.  
 DR MIM; 605459; -.  
 DR MIM; 210250; -.  
 DR GO; GO:0030299; P:cholesterol absorption; NAS.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; FALSE\_NEG.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.



```

Db      300  FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTK 359
Qy      361  DPPGMFGKLGVLRLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNLTGKAVQDR 420
      | ||:| |||||:|||| ||| ||:|||||:|:|:|:| |||:|
Db      360  DSPGVFSKLGVLRLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDR 419
Qy      421  VGLLYQLVGATPYTGMLNAVNLFPMRAVSDQESQDGLYHKWQMLLAYVLHVLPPFSVIAT 480
      ||||| |||||:||||| |||:| |||||:|
Db      420  VGLLYQFVGATPYTGMLNAVNLFVLRVSDQESQDGLYQKWQMLLAYALHVLPPFSVVAT 479
Qy      481  VIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI 540
      :|||||:||||| |||||:||||| |||||:|||||
Db      480  MIFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSVALLSI 539
Qy      541  SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPM 600
      :|:|:|:|:|:| ||: ||||| ||||| ||||| |||: :|
Db      540  AGVLVSGFGLRNIQEMPIPKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPM 599
Qy      601  CAITQGVQFIEKTCPGATSRETANFLILYGFIPALVILGIVIFKVRDY LISR 652
      || |||:||||| ||||| |||||:|:|:|
Db      600  CAFTQGIQFIEKTCPGATSRTMNFILILYSFIPALVILGIVVFKIRDHLISR 651

```

#### RESULT 4

##### ABG8\_RAT

```

ID  ABG8_RAT      STANDARD;      PRT;      694 AA.
AC  P58428; Q8CIQ5; Q923R7;
DT  28-FEB-2003 (Rel. 41, Created)
DT  15-MAR-2004 (Rel. 43, Last sequence update)
DT  15-MAR-2004 (Rel. 43, Last annotation update)
DE  ATP-binding cassette, sub-family G, member 8 (Sterolin-2).
GN  ABCG8.
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX  NCBI_TaxID=10116;
RN  [1]
RP  SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC  STRAIN=Sprague-Dawley;
RX  MEDLINE=21344600; PubMed=11452359;
RA  Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
RA  Ose L., Stalenhoef A.F.H., Mietinnen T., Bjorkhem I., Bruckert E.,
RA  Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,
RA  Patel S.B.;
RT  "Two genes that map to the STSL locus cause sitosterolemia: genomic
RT  structure and spectrum of mutations involving sterolin-1 and
RT  sterolin-2, encoded by ABCG5 and ABCG8, respectively.";
RL  Am. J. Hum. Genet. 69:278-290(2001).
RN  [2]
RP  REVISIONS TO 3-4.
RA  Lu K., Yu H., Lee M.-H., Patel S.B.;
RL  Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
RN  [3]
RP  SEQUENCE FROM N.A. (ISOFORMS 1 AND 3), AND TISSUE SPECIFICITY.
RC  STRAIN=GH, SHR, SHRSP, Sprague-Dawley, Wistar, Wistar Kyoto, and WKA;
RC  TISSUE=Intestine, and Liver;

```

```

RX PubMed=12783625;
RA Yu H., Pandit B., Klett E., Lee M.-H., Lu K., Helou K., Ikeda I.,
RA Egashira N., Sato M., Klein R., Batta A., Salen G., Patel S.B.;
RT "The rat STSL locus: characterization, chromosomal assignment, and
RT genetic variations in sitosterolemic hypertensive rats.";
RL BMC Cardiovasc. Disord. 3:4-4(2003).
CC -!- FUNCTION: Transporter that appears to play an indispensable role
CC in the selective transport of the dietary cholesterol in and out
CC of the enterocytes and in the selective sterol excretion by the
CC liver into bile.
CC -!- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to
CC ABCG5 along a pathway regulating dietary-sterol absorption and
CC excretion (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=3;
CC IsoId=P58428-3; Sequence=Displayed;
CC Name=1;
CC IsoId=P58428-1; Sequence=VSP_008767;
CC Name=2;
CC IsoId=P58428-2; Sequence=VSP_008767, VSP_000054;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Highest expression in liver, with lower levels
CC in small intestine and colon.
CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
CC subfamily.
CC -!- CAUTION: Seems to have a defective ATP-binding region.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF351785; AAK84831.2; -.
DR EMBL; AY145899; AAN64276.1; -.
DR EMBL; AF404109; AAK85393.1; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW Glycoprotein; Transmembrane; Transport; Alternative splicing.
FT DOMAIN 1 434 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 435 455 1 (POTENTIAL).
FT DOMAIN 456 468 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 469 489 2 (POTENTIAL).
FT DOMAIN 490 517 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 518 538 3 (POTENTIAL).
FT DOMAIN 539 547 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 548 568 4 (POTENTIAL).
FT DOMAIN 569 590 CYTOPLASMIC (POTENTIAL).

```

FT	TRANSMEM	591	611	5 (POTENTIAL).
FT	DOMAIN	612	650	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	651	671	6 (POTENTIAL).
FT	DOMAIN	672	694	CYTOPLASMIC (POTENTIAL).
FT	CARBOHYD	640	640	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPPLIC	56	77	Missing (in isoform 1 and isoform 2).
FT				/FTid=VSP_008767.
FT	VARSPPLIC	398	398	Missing (in isoform 2).
FT				/FTid=VSP_000054.
FT	CONFLICT	3	4	EK -> QT (IN REF. 3).
SQ	SEQUENCE	694 AA;	78236 MW;	67F67C195F417587 CRC64;

Query Match 21.4%; Score 720.5; DB 1; Length 694;  
 Best Local Similarity 30.6%; Pred. No. 2.7e-44;  
 Matches 196; Conservative 122; Mismatches 249; Indels 73; Gaps 19;

Qy	20	GSLSSLEQGSVT-----GTEARH-SLGVLHVSYSVSNRVGPW-----WNI	58
		:   :              :	
Db	42	GQSNTLEVRDLTYQGGTCLRSWGQEDPHMSLG-LSESVDMAQV-PWFEQLAQFKLPWRS	99
Qy	59	KSCQQKWDRQILKDVSLYIESGQIMCILGSSSGSKTTLDDAISGRLRRTGTLEGEVFN	118
		:        :   :     :   :       :        :     :	
Db	100	RGSQDSWDLGI-RNLSFKVRSGQMLAIGSAGCGRATLLDVITGRDHGGKMKSGQIWIN	158
Qy	119	CELRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRS-SADFYNNKKEAVMTE	177
		:       :         :   :   :     :       :	
Db	159	QPSTPQLIQKCVAHVRQQDQLLPNLTVRETLTFIAQMRLPKTFSSQAQRDKRVEDVIAELR	218
Qy	178	LSHVADQMIGSYNFGGSISSGERRRVSIAAQLQDPKVMMLDEPTTGLDCMTANQIVLL	237
		:   :                    :   :       :       :	
Db	219	LRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVRTLS	278
Qy	238	ELARRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNCGYPCPEH	297
		:   :   :   :   :   :   :   :   :   :   :   :   :	
Db	279	RLAKGNRLVLISLHQPRSDIFRLFDLVLMTSGTPIYLGVAQHMVQYFTSIGYPCPRYS	338
Qy	298	PFDYMDLTSVDTQSREREIETYKRVQMLECAFKE-----SDIYHKI-LENIERARYL	351
		:     :   :   :   :   :   :   :   :   :   :   :	
Db	339	PADFYVDLTSIDRRSKEQEVATMEKARLLAALFLEKVQGFDDFLWKAEAKSLDTGT	398
Qy	352	LPMVPFKTKDP-----PGMFGKLGVLRRVTRNLNRNKQAVIMRLVQNLIMGLFLIF	403
		:   :       :       :       :       :	
Db	399	QTL----TQDTNCGTAAELPGMIQQFTTLIRRIQISNDFRDLPTLFIHGAEACLSLI	454
Qy	404	YLLRVQNNLTGKAVQDRVGLLYQLVGATPYTGMLNAVNLFPMRLRAVSDQESQDGLYH	463
		:   :      :   :   :   :   :   :   :   :	
Db	455	LYYGHADKPL--SFMDMAALLFMIGALIPFNVLVDVSKCHSERSLLYYELEDGLYTAGP	512
Qy	464	MLLAYVLHVLPPFSVIATVIFSSVCYWTGLYPEVAREFGYFSAALLAPHLIGEF----	517
		:               :	
Db	513	YFFAKVLGELPEHCAYVYIYGMPYIWLTLNLRP-----GPFLFLHFMLLWLWVFC	564
Qy	518	TLVLLGIVQNPNI-VNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEIL	576
		:     :   : :   :   :   :   :   :   :	
Db	565	TMALAASAMLPTEFHMSSFCCNALYNSFYLTAGFMINLNLWIVPAWISKMSFLRWCFSG	624
Qy	577	VVNEFYG-----LNFTCGGSN--TSM-LN-HPMCAI	603

Db                   : :| |           | |: |       |:| || |: ||  
 625 MQIQFNNGHIYTTQIGNLTFVPGDAMVTAMDLSHPLYAI 664

RESULT 5

ABG8\_MOUSE

ID ABG8\_MOUSE           STANDARD;           PRT;    673 AA.  
 AC Q9DBM0;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE ATP-binding cassette, sub-family G, member 8 (Sterolin-2).  
 GN ABCG8.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC STRAIN=C57BL/6; TISSUE=Liver;  
 RX MEDLINE=21344600; PubMed=11452359;  
 RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,  
 RA Ose L., Stalenhoef A.F.H., Mietinnen T., Bjorkhem I., Bruckert E.,  
 RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,  
 RA Patel S.B.;  
 RT "Two genes that map to the STSL locus cause sitosterolemia: genomic  
 RT structure and spectrum of mutations involving sterolin-1 and  
 RT sterolin-2, encoded by ABCG5 and ABCG8, respectively."  
 RL Am. J. Hum. Genet. 69:278-290(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC STRAIN=C57BL/6J; TISSUE=Liver;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP TISSUE SPECIFICITY, AND INDUCTION.  
 RX MEDLINE=20553648; PubMed=11099417;  
 RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,

RA Kwiterovich P., Shan B., Barnes R., Hobbs H.H.;  
 RT "Accumulation of dietary cholesterol in sitosterolemia caused by  
 RT mutations in adjacent ABC transporters.";  
 RL Science 290:1771-1775(2000).  
 CC -!- FUNCTION: Transporter that appears to play an indispensable role  
 CC in the selective transport of the dietary cholesterol in and out  
 CC of the enterocytes and in the selective sterol excretion by the  
 CC liver into bile.  
 CC -!- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to  
 CC ABCG5 along a pathway regulating dietary-sterol absorption and  
 CC excretion (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q9DBM0-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q9DBM0-2; Sequence=VSP\_000053;  
 CC Note=No experimental confirmation available;  
 CC -!- TISSUE SPECIFICITY: Expressed in the intestine and, at lower  
 CC level, in the liver.  
 CC -!- INDUCTION: Upregulated by cholesterol feeding. Possibly mediated  
 CC by the liver X receptor/retinoide X receptor (LXR/RXR) pathway.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
 CC subfamily.  
 CC -!- CAUTION: Seems to have a defective ATP-binding region.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF324495; AAK84079.1; -.  
 DR EMBL; AK004871; BAB23630.1; -.  
 DR MGD; MGI:1914720; Abcg8.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 KW Glycoprotein; Transmembrane; Transport; Alternative splicing.  
 FT DOMAIN 1 413 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 414 434 1 (POTENTIAL).  
 FT DOMAIN 435 447 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 448 468 2 (POTENTIAL).  
 FT DOMAIN 469 496 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 497 517 3 (POTENTIAL).  
 FT DOMAIN 518 526 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 527 547 4 (POTENTIAL).  
 FT DOMAIN 548 569 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 570 590 5 (POTENTIAL).  
 FT DOMAIN 591 639 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 640 660 6 (POTENTIAL).  
 FT DOMAIN 661 673 CYTOPLASMIC (POTENTIAL).

FT	CARBOHYD	619	619	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSP LIC	377	377	Missing (in isoform 2).
FT				/FTId=VSP_000053.
SQ	SEQUENCE	673 AA;	75995 MW;	78012611A5DF2589 CRC64;

[illegible]



# RESULT 6

## ABG2\_HUMAN

ID ABG2\_HUMAN STANDARD; PRT; 655 AA.  
AC Q9UNQ0; O95374; Q9BY73; Q9NUS0;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE ATP-binding cassette, sub-family G, member 2 (Placenta-specific ATP-binding cassette transporter) (Breast cancer resistance protein).  
GN ABCG2 OR ABCP OR BCRP OR BCRP1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=99065313; PubMed=9850061;  
RA Allikmets R., Schriml L.M., Hutchinson A., Romano-Spica V., Dean M.;  
RT "A human placenta-specific ATP-binding cassette gene (ABCP) on  
RT chromosome 4q22 that is involved in multidrug resistance.";  
RL Cancer Res. 58:5337-5339(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Breast cancer;  
RX MEDLINE=99080071; PubMed=9861027;  
RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,  
RA Ross D.D.;  
RT "A multidrug resistance transporter from human MCF-7 breast cancer  
RT cells.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:15665-15670(1998).  
RN [3]  
RP ERRATUM.  
RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,  
RA Ross D.D.;  
RL Proc. Natl. Acad. Sci. U.S.A. 96:2569-2569(1999).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Kage K., Tsukahara S., Sugiyama T., Asada S., Ishikawa E., Tsuruo T.,  
RA Sugimoto Y.;  
RT "Breast cancer resistance protein constitutes a 140-kDa complex as a  
RT homodimer.";  
RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.  
RN [5]  
RP SEQUENCE OF 198-655 FROM N.A.  
RC TISSUE=Placenta;  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,  
RA Ninomiya K., Iwayanagi T.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.  
RN [6]

RP REVIEW.  
RX MEDLINE=21474438; PubMed=11590207;  
RA Schmitz G., Langmann T., Heimerl S.;  
RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";  
RL J. Lipid Res. 42:1513-1520(2001).  
CC -!- FUNCTION: Xenobiotic transporter that appears to play a major role  
CC in the multidrug resistance phenotype of a specific MCF-7 breast  
CC cancer cell line. When overexpressed, the transfected cells become  
CC resistant to mitoxantrone, daunorubicin and doxorubicin, display  
CC diminished intracellular accumulation of daunorubicin, and  
CC manifest an ATP-dependent increase in the efflux of rhodamine 123.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
CC subfamily.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AF103796; AAD09188.1; -.  
DR EMBL; AF098951; AAC97367.1; -.  
DR EMBL; AB056867; BAB39212.1; -.  
DR EMBL; AK002040; BAA92050.1; -.  
DR Genew; HGNC:74; ABCG2.  
DR MIM; 603756; -.  
DR GO; GO:0016021; C:integral to membrane; TAS.  
DR GO; GO:0005524; F:ATP binding; TAS.  
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; TAS.  
DR GO; GO:0005215; F:transporter activity; TAS.  
DR GO; GO:0008559; F:xenobiotic-transporting ATPase activity; TAS.  
DR GO; GO:0009315; P:drug resistance; TAS.  
DR GO; GO:0006810; P:transport; TAS.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003439; ABC\_transporter.  
DR Pfam; PF00005; ABC\_tran; 1.  
DR ProDom; PD000006; ABC\_transporter; 1.  
DR SMART; SM00382; AAA; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; FALSE\_NEG.  
DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
KW ATP-binding; Transmembrane; Transport.

FT	DOMAIN	1	395	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	396	416	POTENTIAL.
FT	DOMAIN	417	428	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	429	449	POTENTIAL.
FT	DOMAIN	450	477	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	478	498	POTENTIAL.
FT	DOMAIN	499	506	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	507	527	POTENTIAL.
FT	DOMAIN	528	535	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	536	556	POTENTIAL.
FT	DOMAIN	557	630	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	631	651	POTENTIAL.
FT	DOMAIN	652	655	CYTOPLASMIC (POTENTIAL).



## RESULT 7

## ABG8\_HUMAN

ID ABG8\_HUMAN STANDARD; PRT; 673 AA.  
AC Q9H221;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE ATP-binding cassette, sub-family G, member 8 (Sterolin-2).  
GN ABCG8.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., VARIANTS SITOSTEROLEMIA THR-231; GLN-263; ARG-574  
RP AND ARG-596, AND VARIANT CYS-54.  
RX MEDLINE=20553648; PubMed=11099417;  
RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,  
RA Kwiterovich P., Shan B., Barnes R., Hobbs H.H.;  
RT "Accumulation of dietary cholesterol in sitosterolemia caused by  
RT mutations in adjacent ABC transporters.";  
RL Science 290:1771-1775(2000).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), VARIANTS SITOSTEROLEMIA  
RP HIS-184; THR-231; GLN-263; HIS-405; PRO-501; SER-543; PRO-572;  
RP GLU-574; ARG-574; ARG-596 AND PHE-570 DEL, AND VARIANTS HIS-19;  
RP CYS-54; LYS-238; VAL-259; LYS-400; ARG-575 AND ALA-632.  
RC TISSUE=Liver;  
RX MEDLINE=21344600; PubMed=11452359;  
RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,  
RA Ose L., Stalenhoef A.F.H., Mietinnen T., Bjorkhem I., Bruckert E.,  
RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,  
RA Patel S.B.;  
RT "Two genes that map to the STSL locus cause sitosterolemia: genomic  
RT structure and spectrum of mutations involving sterolin-1 and  
RT sterolin-2, encoded by ABCG5 and ABCG8, respectively.";  
RL Am. J. Hum. Genet. 69:278-290(2001).  
RN [3]  
RP REVIEW.  
RX MEDLINE=21474438; PubMed=11590207;  
RA Schmitz G., Langmann T., Heimerl S.;  
RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";  
RL J. Lipid Res. 42:1513-1520(2001).  
CC -!- FUNCTION: Transporter that appears to play an indispensable role  
CC in the selective transport of the dietary cholesterol in and out  
CC of the enterocytes and in the selective sterol excretion by the  
CC liver into bile.  
CC -!- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to  
CC ABCG5 along a pathway regulating dietary-sterol absorption and  
CC excretion.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1;

```

CC      IsoId=Q9H221-1; Sequence=Displayed;
CC      Name=2;
CC      IsoId=Q9H221-2; Sequence=VSP_000052;
CC      Note=Minor form detected in approximately 10% of the cDNA
CC      clones;
CC      -!- TISSUE SPECIFICITY: Strongly expressed in the liver, lower levels
CC          in the small intestine and colon. Detectable in a wide variety of
CC          human tissues.
CC      -!- DISEASE: Defects in ABCG8 are a cause of sitosterolemia
CC          [MIM:210250]; also known as phytosterolemia or shellfish
CC          sterolemia. It is a rare autosomal recessive disorder
CC          characterized by increased intestinal absorption of all sterols
CC          including cholesterol, plant and shellfish sterols, and decreased
CC          biliary excretion of dietary sterols into bile. Sitosterolemia
CC          patients have hypercholesterolemia, very high levels of plant
CC          sterols in the plasma, and frequently develop tendon and tuberous
CC          xanthomas, accelerated atherosclerosis and premature coronary
CC          artery disease.
CC      -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
CC          subfamily.
CC      -!- CAUTION: Seems to have a defective ATP-binding region.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AF320294; AAG40004.1; -.
DR      EMBL; AF324494; AAK84078.1; -.
DR      EMBL; AF351824; AAK84663.1; -.
DR      EMBL; AF351812; AAK84663.1; JOINED.
DR      EMBL; AF351813; AAK84663.1; JOINED.
DR      EMBL; AF351814; AAK84663.1; JOINED.
DR      EMBL; AF351815; AAK84663.1; JOINED.
DR      EMBL; AF351816; AAK84663.1; JOINED.
DR      EMBL; AF351817; AAK84663.1; JOINED.
DR      EMBL; AF351818; AAK84663.1; JOINED.
DR      EMBL; AF351819; AAK84663.1; JOINED.
DR      EMBL; AF351820; AAK84663.1; JOINED.
DR      EMBL; AF351821; AAK84663.1; JOINED.
DR      EMBL; AF351822; AAK84663.1; JOINED.
DR      EMBL; AF351823; AAK84663.1; JOINED.
DR      Genew; HGNC:13887; ABCG8.
DR      MIM; 605460; -.
DR      MIM; 210250; -.
DR      InterPro; IPR003439; ABC_transporter.
DR      Pfam; PF00005; ABC_tran; 1.
DR      ProDom; PD000006; ABC_transporter; 1.
DR      PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR      PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW      Glycoprotein; Transmembrane; Transport; Alternative splicing;
KW      Polymorphism; Disease mutation.
FT      DOMAIN      1      416      CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM    417    437      1 (POTENTIAL).

```

FT	DOMAIN	438	447	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	448	468	2 (POTENTIAL).
FT	DOMAIN	469	492	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	493	513	3 (POTENTIAL).
FT	DOMAIN	514	531	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	532	552	4 (POTENTIAL).
FT	DOMAIN	553	569	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	570	590	5 (POTENTIAL).
FT	DOMAIN	591	639	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	640	660	6 (POTENTIAL).
FT	DOMAIN	661	673	CYTOPLASMIC (POTENTIAL).
FT	CARBOHYD	619	619	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	376	376	Missing (in isoform 2).
FT				/FTId=VSP_000052.
FT	VARIANT	19	19	D -> H.
FT				/FTId=VAR_012250.
FT	VARIANT	54	54	Y -> C.
FT				/FTId=VAR_012251.
FT	VARIANT	184	184	R -> H (in sitosterolemia).
FT				/FTId=VAR_012252.
FT	VARIANT	231	231	P -> T (in sitosterolemia).
FT				/FTId=VAR_012253.
FT	VARIANT	238	238	E -> K.
FT				/FTId=VAR_012254.
FT	VARIANT	259	259	A -> V.
FT				/FTId=VAR_012255.
FT	VARIANT	263	263	R -> Q (in sitosterolemia).
FT				/FTId=VAR_012256.
FT	VARIANT	400	400	T -> K.
FT				/FTId=VAR_012257.
FT	VARIANT	405	405	R -> H (in sitosterolemia).
FT				/FTId=VAR_012258.
FT	VARIANT	501	501	L -> P (in sitosterolemia).
FT				/FTId=VAR_012259.
FT	VARIANT	543	543	R -> S (in sitosterolemia).
FT				/FTId=VAR_012260.
FT	VARIANT	570	570	Missing (in sitosterolemia).
FT				/FTId=VAR_012261.
FT	VARIANT	572	572	L -> P (in sitosterolemia).
FT				/FTId=VAR_012262.
FT	VARIANT	574	574	G -> E (in sitosterolemia).
FT				/FTId=VAR_012263.
FT	VARIANT	574	574	G -> R (in sitosterolemia).
FT				/FTId=VAR_012264.
FT	VARIANT	575	575	G -> R.
FT				/FTId=VAR_012265.
FT	VARIANT	596	596	L -> R (in sitosterolemia).
FT				/FTId=VAR_012266.
FT	VARIANT	632	632	V -> A.
FT				/FTId=VAR_012267.
SQ	SEQUENCE	673 AA;	75678 MW;	594AFD1D6C1BB50F CRC64;

Query Match 20.4%; Score 688.5; DB 1; Length 673;  
 Best Local Similarity 28.1%; Pred. No. 5.2e-42;  
 Matches 188; Conservative 125; Mismatches 233; Indels 123; Gaps 16;

Qy 37 HSLGVLHVSYSV--SNRVGPW-----WNIKSCQKQWDRQILKDVSLYIESGQIMC 84

```

Db      45 NTLEVRDLNYQVDLASQV-PWFEQLAQFKMPWTSPSCQNSCELGI-QNLSFKVRSGQMLA 102
Qy      85 ILGSSGSGKTLLDAISGRLRRTGTLEGEVFNVCCELRRDQFQDCFSYVLQSDVFLSSLT 144
      |:|||| |: :||| |:|| |:::| | :| :| :| :|
Db     103 IIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQPPSSPQLVRKCVAHVRQHNQLLPNLT 162
Qy     145 VRETLRYTAMLALCRS-SADFYNNKKVEAVMTELSLSHVADQMIGSYNFGGISSGERRRVS 203
      ||||| :| :| |:| :|:| |:|| | | :| :| :| :| :| :|
Db     163 VRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVS 222
Qy     204 IAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVVTIHQPRSELFQHFQDK 263
      | ||| :| :|:| ||| ||: :| |: ||: |:|:|:|:|:|:|:|
Db     223 IGVQLLWNPGLILDEPTSGLDSTAHNLVKTLRLAKGNRLVLISLHQPRSDIFRLFDL 282
Qy     264 IAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRV 323
      : :| | :| :| :| :| |||| :|| |||:|:|:| :|:|:| :| :|
Db     283 VLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKA 342
Qy     324 QMLECAFKESDIYHKILENIERARYL-----KTLPM----VPFKT 359
      | | | :|:| | | :| :| |
Db     343 QSLAALF-----LEKVRDLDDFLWKAETKDLDEDTCESSVTPLDTNCLPSPT 390
Qy     360 KDPPGMFGKLGVLRLRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNTLKGAVQD 419
      | || :| :| | |: :| :| :| :| :| :| :| :| :|
Db     391 K-MPGAVQQFTTLIRRQISNDFRDLPTLLIHGAECILMSMTIGF--LYFGHGSIQLSFMD 447
Qy     420 RVGLLYQLVGATPYTGMLNAVNLFPMRAVSDQESQDGLYHKWQMLLAYVLHVLFPFSVIA 479
      ||: :| |: :|: :| ||: | :||| | :| ||
Db     448 TAALLFMIGALIPFNVILDVISKYSERAMLYYELEDGLYTTGPYFFAKILGELPEHCAY 507
Qy     480 TVIFSSVCYWTGLGYPEVARF-----GYFSAALLAPHLIGEFLLTVLL 522
      :|: || | | :| | :||| :| :| |
Db     508 IIIYGMPTYWLANLRPGLQPFLLHFLVLVVFCCRIMALAAAALLPTFHMASFFSNAL- 566
Qy     523 GIVQNPNIVNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEF- 581
      : : ||: |: : : :| :| | |: :|
Db     567 -----YNSFYLAGGFMINLSSLWTVPAWISKVSFLRWCFEGLMKIQFS 609
Qy     582 ---YGL---NFTCGGSNTSML-----NHPMCAITQGVQFIEKTCPGATSRFTANFLILY 629
      | : | | | :| :|:| | :||:
Db     610 RRTYKMPLGNLTIAVSGDKILSVMELDSYPLYAI-----YLIVI 648
Qy     630 GFIPALVIL 638
      | :|
Db     649 GLSGGFMVL 657

```

RESULT 8

YOH5\_YEAST

ID YOH5\_YEAST STANDARD; PRT; 1294 AA.

AC Q08234; Q08233;

DT 01-NOV-1997 (Rel. 35, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Probable ATP-dependent transporter YOL074C/YOL075C.

GN YOL074C/YOL075C.

OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97321807; PubMed=9178509;  
 RA Tzermia M., Katsoulou C., Alexandraki D.;  
 RT "Sequence analysis of a 33.2 kb segment from the left arm of yeast  
 RT chromosome XV reveals eight known genes and ten new open reading  
 RT frames including homologues of ABC transporters, inositol  
 RT phosphatases and human expressed sequence tags.";  
 RL Yeast 13:583-589(1997).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z74817; CAA99085.1; -.  
 DR EMBL; Z74816; CAA99084.1; -.  
 DR PIR; S77690; S77690.  
 DR GermOnline; 143497; -.  
 DR SGD; S0005435; YOL075C.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 2.  
 DR ProDom; PD000006; ABC\_transporter; 2.  
 DR SMART; SM00382; AAA; 2.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 2.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 2.  
 KW Hypothetical protein; ATP-binding; Transmembrane; Glycoprotein;  
 KW Transport; Repeat.  
 FT TRANSMEM 376 396 POTENTIAL.  
 FT TRANSMEM 496 516 POTENTIAL.  
 FT TRANSMEM 531 551 POTENTIAL.  
 FT TRANSMEM 605 625 POTENTIAL.  
 FT TRANSMEM 1039 1059 POTENTIAL.  
 FT TRANSMEM 1121 1141 POTENTIAL.  
 FT TRANSMEM 1267 1287 POTENTIAL.  
 FT NP\_BIND 62 69 ATP (POTENTIAL).  
 FT NP\_BIND 727 734 ATP (POTENTIAL).  
 FT CARBOHYD 41 41 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 341 341 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 349 349 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 371 371 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 983 983 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1062 1062 N-LINKED (GLCNAC. . .) (POTENTIAL).



SQ SEQUENCE 1294 AA; 145157 MW; C555500A45E9284E CRC64;

Query Match 18.6%; Score 628; DB 1; Length 1294;  
Best Local Similarity 29.6%; Pred. No. 2.8e-37;  
Matches 183; Conservative 123; Mismatches 237; Indels 76; Gaps 21;

```
Qy      67 RQILKDVSLYIESGQIMCILGSSSGSKTTLDDAISGRLRRT---GTLEGEVVFVNGCELR 122
      ::||: |: : | | |:| ||||:|:|: ||||: : : | : | ::
Db      707 KEILQSVNAIFKPGMINAIMGPSGSGKSSLLNLISGRLKSSVFAKFDTSGSIMFNDIQVS 766

Qy      123 RDQFQDCFSYVLQ-SDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTELSLSHV 181
      |:: ||| | | |:||||:|:| | | | : : : :: | | |
Db      767 ELMFKNVCSYVSQDDHLLAALTVKETLKYAAALRLHHLTEAERMERTDNLIRSLGLKHC 826

Qy      182 ADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELAR 241
      : :||: ||| ||:|:|: || | | : :|||:| | | |: |: :| :| |
Db      827 ENNIIGNEFVKGISGGEKRRVTMGVQLLNDPPILLDEPTSGLDSFTSATILEILEKLCR 886

Qy      242 -RDRIVIVTIHQPRSELFQHFQDKIALIT-YGELVFCGTPEEMLGFFNNCGYPCPEHSNPF 299
      : : :|:|||||:|:| : : | | | |:|:|: :| | | | :|
Db      887 EQGKTIITITIHQPRSELFKRFGNVLLAKSGRTAFNGSPDEMIAYFTELGYNCPSFTNVA 946

Qy      300 DFYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILEN-----IERARYLKT 351
      ||:| | |:|: : | | : | : : | : : :| : : | :
Db      947 DFFLDLISVNTQNEQNEISSRARVEKILSAWKA-----MDNESLSPTPISEKQQYSQE 1000

Qy      352 LPMVPFK--TKDPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQ 409
      : : : | : | : | : | : | : | : | : | : | : | :
Db      1001 SFFTEYSEFVRKPANLVLAYIVNVKRQFTTTRRSFDSLARIAQIPGLGVIFALFFAPVK 1060

Qy      410 NNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYV 469
      :| : : :|:| : | : | | : : | | : | : | : | : | :
Db      1061 HNYT--SISNRLGLAQEST-ALYFVGMLGNLACYPTERDYFYEEYNDNVYGIAPFFFLAYM 1117

Qy      470 LHVLPFSVIATVIFSSVCYWTGLGYPEVARFGYFSAALLAPHLI---GEFLTLLVLLGIVQ 526
      || | :|:|: : || | | | | : : : || | : : :
Db      1118 TLELPLSALASVLYAVFTVLACGL-PRTA--GNFFATVYCSFIVTCCGERLGIMTNTFFE 1174

Qy      527 NPN-IVNSIVALLSI---SGLL-IGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNE 580
      | :| | :| | ||: :| | | | | | | :| :
Db      1175 RPGFVWNCISIILSIGTQMSGLMSLG-----MSRVLKGFNYLNPVGYTSMIINF 1225

Qy      581 FYG-LNFTC--GGSNTSMLNHPMCAITQGVQFIEKTCPGATSREFTANFLILYGFIP---- 633
      | | | || || |: : || | : |: | :
Db      1226 FPGNLKLTCEDGGKNS-----DGTCEFANGH---DVLVSYGLVRNTQK 1265

Qy      634 --ALVILGIVIFKVRDYLI 650
      : : : |: : : : |
Db      1266 YLGIIVCVAIYRLIAFFI 1284
```

RESULT 9

ADP1\_YEAST

ID ADP1\_YEAST STANDARD; PRT; 1049 AA.

AC P25371;

DT 01-MAY-1992 (Rel. 22, Created)

DT 01-MAY-1992 (Rel. 22, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable ATP-dependent permease precursor.  
 GN ADP1 OR YCR011C OR YCR11C OR YCR105.  
 OS *Saccharomyces cerevisiae* (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; *Saccharomyces*.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92160395; PubMed=1789009;  
 RA Purnelle B., Skala J., Goffeau A.;  
 RT "The product of the YCR105 gene located on the chromosome III from  
 RT *Saccharomyces cerevisiae* presents homologies to ATP-dependent  
 RT permeases.";  
 RL Yeast 7:867-872(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92327849; PubMed=1626432;  
 RA Skala J., Purnelle B., Goffeau A.;  
 RT "The complete sequence of a 10.8 kb segment distal of *SUF2* on the  
 RT right arm of chromosome III from *Saccharomyces cerevisiae* reveals  
 RT seven open reading frames including the *RVS161*, *ADP1* and *PGK* genes.";  
 RL Yeast 8:409-417(1992).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X59720; CAA42328.1; -.  
 DR PIR; S19421; S19421.  
 DR GermOnline; 138916; -.  
 DR SGD; S0000604; ADP1.  
 DR GO; GO:0005783; C:endoplasmic reticulum; IDA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 KW ATP-binding; Transmembrane; Glycoprotein; Transport; Signal.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 1049 PROBABLE ATP-DEPENDENT PERMEASE.  
 FT NP\_BIND 423 430 ATP (BY SIMILARITY).  
 FT TRANSMEM 325 345 POTENTIAL.  
 FT TRANSMEM 464 481 POTENTIAL.  
 FT TRANSMEM 794 814 POTENTIAL.  
 FT TRANSMEM 829 849 POTENTIAL.  
 FT TRANSMEM 878 898 POTENTIAL.  
 FT TRANSMEM 910 930 POTENTIAL.  
 FT TRANSMEM 938 958 POTENTIAL.

FT	TRANSMEM	1001	1021	POTENTIAL.
FT	TRANSMEM	1025	1045	POTENTIAL.
FT	CARBOHYD	50	50	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	114	114	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	165	165	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	221	221	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	815	815	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	935	935	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	960	960	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	971	971	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	1049 AA;	117231 MW;	ABC9CE54BCFDF6A3 CRC64;

Query Match 17.9%; Score 602.5; DB 1; Length 1049;  
 Best Local Similarity 26.5%; Pred. No. 1.5e-35;  
 Matches 191; Conservative 130; Mismatches 227; Indels 173; Gaps 25;

Qy	38	SLGVLHVSYSVSNRVGPWWNIKSCQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLL	97
		:  ::        :: :  ::  ::     :      :	
Db	383	TLSFENITYSV-----PSINSDGVEE-----TVLNEISGIVKPGQILAIMGGSGAGKTTLL	433
Qy	98	DAISGRLRRTGTLEGEVFVNGCELRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLAL	157
		:: :  ::  :   :     :     :        :      : : :	
Db	434	DILAMK-RKTGHVSGSIKVNIGISMDRKSFSKIIGFVDQDDFLLPTLTVFETVLNSALLRL	492
Qy	158	CRS-SADFYNNKKVEAVMTELSLSHVADQMIGSYNFGGISSGERRRVSIQAQLLQDPKVM	216
		::   : :   ::    : :  ::  :       :      : :    :	
Db	493	PKALSFEAKKARVYKVLEELRIIDIKDRIIGNEFDRGISGGEKRRVSIACELVTSPLVLF	552
Qy	217	LDEPTTGLDCMTANQIVLLLAELAR-RDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFC	275
		:        ::     : :  ::::       :      : :    : :	
Db	553	LDEPTSGLDASNANNVIECLVRLSSDYNRTLVLSTHQPNSNIFYLFDKLVLLSKGEMVYS	612
Qy	276	GTPEEMLGFFNNCGYPCPEHSNPFDFYMDLT-----SVD-	309
		:: :          :     : : :  :	
Db	613	GNAKKVSEFLRNEGICPDNYNIADYLIDITFEAGPQGKRRRIRNISDLEAGTDTNDIDN	672
Qy	310	-----TQSREREIETYKRVQMLECA	329
		:	
Db	673	TIHQTTFTSSDGTQREWAHLAAHRDEIRSLLRDEEDVEGTDGRRGATEIDLNTKLLHDK	732
Qy	330	FKESDIYHKILENI-----ERARYLK-TLPMVPFKTKDPPGMFGKLGVLRLRVTRNL	380
		: :    :: :     :      :   :  :    : :	
Db	733	YKDSVYYAELSQEIEEVLSEGDEESNVLNGDLP-----TGQQSAGFLQQLSILNSRSFKNM	788
Qy	381	MRNQAVIMRLVQNLMGLFL--IFYLLRVQNNTLKGAVQDRVGLLYQLV---GATPYTG	435
		: :: : ::     ::  : :  :    ::   : ::   :	
Db	789	YRNPKLLLGNYLLTILLSLFLGTLYYNV---SNDISG-FQNRMGLEFFILTYFGFVTFTG	844
Qy	436	MLNAVNLFPMRLRAVSDQESQDGLYHKWQMLLAYVL-----HVLPFSVIATVIFSSVCYWT	490
		: :   :   : :  :       :   :    : :  :	
Db	845	L----SSFALERIIFIKERSNNYYSP----LAYYISKIMSEVVPLRVVPPILLSLIVYPM	896
Qy	491	LGLYPEVARFGYFSAALLAPHLIGEFLLTLVLLGIV---QNPNIIVNSIVALLS---ISGLL	544
		:   : :  :    :: :  :   :  : :	
Db	897	TGLNMKDNAF-FKCIGILILFNLGISLEILTIGIIFEDLNNIILSVLVLLGSLFSGLF	955
Qy	545	IGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEF-----YGLNFTCGGSNTSM	595

```

          | : | |      : | | | : | | | : | |
Db      956 INTKNITN-----VAFKYLKNFVSFYYAYESLLINEVKTMLMKERKYGLNI----- 1001

Qy      596 LNHPMCAITQGVQFIEKTCPGATS RFTANFLILYGFIPALVILGI-----VIFKVRDY 648
          | | | |      | : | | : : |      | : | : |
Db      1002 -----EVPGAT-----ILSTFGFVVQNLVFDIKILALFNVVFLIMGY 1038

Qy      649 L 649
          |
Db      1039 L 1039

```

# RESULT 10

## WHIT\_DROME

```

ID  WHIT_DROME      STANDARD;      PRT;      687 AA.
AC  P10090; Q9V3A2; Q9XY33;
DT  01-MAR-1989 (Rel. 10, Created)
DT  01-NOV-1991 (Rel. 20, Last sequence update)
DT  10-OCT-2003 (Rel. 42, Last annotation update)
DE  White protein.
GN  W OR EG:BACN33B1.1 OR CG2759.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Head;
RX  MEDLINE=90221897; PubMed=2109311;
RA  Pepling M., Mount S.M.;
RT  "Sequence of a cDNA from the Drosophila melanogaster white gene.";
RL  Nucleic Acids Res. 18:1633-1633(1990).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=85134865; PubMed=6084717;
RA  O'Hare K., Murphy C., Levis R., Rubin G.M.;
RT  "DNA sequence of the white locus of Drosophila melanogaster.";
RL  J. Mol. Biol. 180:437-455(1984).
RN  [3]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=21100348; PubMed=11156992;
RA  Lukacsovich T., Asztalos Z., Awano W., Baba K., Kondo S., Niwa S.,
RA  Yamamoto D.;
RT  "Dual-tagging gene trap of novel genes in Drosophila melanogaster.";
RL  Genetics 157:727-742(2001).
RN  [4]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Berkeley;
RX  MEDLINE=20196006; PubMed=10731132;
RA  Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA  Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA  George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA  Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA  Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA  Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA  Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

```

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Oregon-R;  
 RX MEDLINE=20196011; PubMed=10731137;  
 RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,  
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,  
 RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,  
 RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,  
 RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,  
 RA Modolell J., Peter A., Schoettler P., Werner M., Mourkioti F.,  
 RA Beinert N., Dowe G., Schaefer U., Jaeckle H., Bucheton A.,  
 RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,  
 RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,  
 RA Glover D.M.;  
 RT "From sequence to chromosome: the tip of the X chromosome of *D.*  
 RT *melanogaster*.";  
 RL Science 287:2220-2222(2000).  
 RN [6]  
 RP SEQUENCE OF 224-331 FROM N.A.  
 RX MEDLINE=89339145; PubMed=2503416;  
 RA Tearle R.G., Belote J.M., McKeown M., Baker B.S., Howells A.J.;  
 RT "Cloning and characterization of the scarlet gene of *Drosophila*  
 RT *melanogaster*.";  
 RL Genetics 122:595-606(1989).  
 CC -!- FUNCTION: Part of a membrane-spanning permease system necessary  
 CC for the transport of pigment precursors into pigment cells

CC responsible for eye color. White dimerize with brown for the  
 CC transport of guanine and with scarlet for the transport of  
 CC tryptophan.  
 CC -!- SUBUNIT: Heterodimer of white with either brown or scarlet.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X51749; CAA36038.1; -.  
 DR EMBL; X02974; CAA26716.1; -.  
 DR EMBL; AB028139; BAA78210.1; -.  
 DR EMBL; AE003425; AAF45826.1; -.  
 DR EMBL; AL133506; CAB65847.1; -.  
 DR EMBL; X76202; CAA53795.1; -.  
 DR PIR; S08635; FYFFW.  
 DR FlyBase; FBgn0003996; w.  
 DR GO; GO:0004888; F:transmembrane receptor activity; NAS.  
 DR GO; GO:0006727; P:ommochrome biosynthesis; IMP.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR005284; Pigment\_permease.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGRFAMs; TIGR00955; 3a01204; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 KW Pigment; ATP-binding; Transmembrane; Transport.  
 FT NP\_BIND 130 137 ATP (BY SIMILARITY).  
 FT TRANSMEM 435 453 POTENTIAL.  
 FT TRANSMEM 465 485 POTENTIAL.  
 FT TRANSMEM 515 533 POTENTIAL.  
 FT TRANSMEM 542 563 POTENTIAL.  
 FT TRANSMEM 576 594 POTENTIAL.  
 FT TRANSMEM 659 678 POTENTIAL.  
 FT CONFLICT 25 29 GDSGA -> LIFEIPYHCRVTAD (IN REF. 2 AND  
 FT 3).  
 FT CONFLICT 49 49 L -> R (IN REF. 4 AND 5).  
 FT CONFLICT 335 371 VGAQCPTNYPADFYVQVLAVVPGREIESRDRIAKIC ->  
 FT ITLHLNSYPWVPSVLPTTIRRTFTYRCWPLCPDGRSSPVI  
 FT GSPRYG (IN REF. 3).  
 SQ SEQUENCE 687 AA; 75672 MW; 24AFAD799DE0D396 CRC64;

Query Match 17.5%; Score 589; DB 1; Length 687;  
 Best Local Similarity 27.5%; Pred. No. 8e-35;  
 Matches 200; Conservative 129; Mismatches 255; Indels 142; Gaps 25;

Qy 11 GARGP---HINRGSLSSLEQ-----GSVTG-----TEAR 36  
 |::| |::| :| |::| |  
 Db 13 GSKHPSAEHLNNGDSGAASQSCINQGFGQAKNYGTLLPPSPPEDSGSGGQLAENLTYAW 72

Qy 37 HSLGVLHVSYSVSNRVGPWWNI-----KSCQQKW----DRQILKDVSLYIESGQIMCI 85  
 |:: : : | : | | : : : : : | : : :  
 Db 73 HNMDI----FGAVNQPGSGWRQLVNRTGLFCNERHIPAPRKHLLKNVCGVAYPGELLAV 128  
 Qy 86 LGSSSGSGKTTLLDAISGRLLR--TGTLEGEVVFVNGCELRRDQFQDCFSYVLQSDVFLSSL 143  
 :|||:|||||:|:: | : : | :|| : : | :|| | |:: ||  
 Db 129 MGSSGAGKTTLLNALAFRSPQGIQVSPSGMRLNGQPVDAKEMQARCAIVQDDLFIGSL 188  
 Qy 144 TVRETLRYTAMLALCRSSADFYNK---KVEAVMTELSLSHVADQMIG-SYNFGGISSGER 199  
 | || | : ||: : | | : :|| : |||| :|| |:: |||  
 Db 189 TAREHLIFQAMVRMPRHLLT--YRQVARVDQVIQELSLSKCQHTIIGVPRVKGLSGGER 246  
 Qy 200 RRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQ 259  
 :||: |:: | || :|| | |||: || |:: :| :| :|::: : ||: |||| | ||:  
 Db 247 KRLAFASEALTDPLLICDEPTSGLDSFTAHSVVQVLKKLSQKGKTVILTIHQPSSEFE 306  
 Qy 260 HFDKIAILTYGELVFCGTPEEMLGFFNNGCGYPCPEHSNPFDFYMDLTSVDTQSREREIET 319  
 |||| : : | : | || | : ||: | || : || |||: : : | |||:  
 Db 307 LFDKILLMAEGRVAFGLTGPSEAVDFFSYVGAQCPTNYPADFYVQVLAV---VPGREIES 363  
 Qy 320 YKRVQMLECAFKESDIYHKILENIERARYLKLTPMVPFKTKDPP-----GMFGKLG 371  
 |: : | | | : :||: | | | : | | : | :  
 Db 364 RDRIAKICDNFAIS----KVARDEQLLATKNLE----KPLEQPENGYTYKATWFMQFRA 415  
 Qy 372 LLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNLTKGAVQDRVGLLYQLVGAT 431  
 :| | :||: | :||: || : : || : | | | | : | : :  
 Db 416 VLWRSWLSVLKEPLLKVRLIQTTMVAI-LIGLIFLGQQLTQVG-VMNINGAIFLFLTNM 473  
 Qy 432 PYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYVLHVLPPFSVIATVIFSSVCYWT 491  
 : : :||: | :||: || | : || : || : :||: : | :  
 Db 474 TFQNVFATINVFTSELFPVFMREARSRLYRCDTYFLGKTIAELPLFLTVPPLVFTAIAYP 533  
 Qy 492 GLYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNI VNSIVALLS----- 539  
 || | | | | | | : : | : | : | :  
 Db 534 GLRAGVLHF-----FNCLALVTLV--ANVSTSGYLISCASSSTSMALSV 576  
 Qy 540 ----ISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGL---NFTCGGSN 592  
 | | : || | :||: || | : : | | | :||: : : | ||  
 Db 577 GPPVIIPFLLFGGFFLNSGSVPVYLKWLSYLSWFRYANEGLLINQWADVEPGEISCTSSN 636  
 Qy 593 TSMLNHMPCAITQGVQFIEKTCPGA-----TSRFTANFLILYGFIPALVILGIVIFKVR 646  
 | | | : | | : | | : | || || | :  
 Db 637 T-----TCPSSGKVILETLNFSAADLPL-DYV-GLAIL-IVSFRVL 674  
 Qy 647 DYLI SR 652  
 || |  
 Db 675 AYLA LR 680

RESULT 11

ABG1\_HUMAN

ID ABG1\_HUMAN STANDARD; PRT; 678 AA.

AC P45844; Q9BXX6; Q9BXX7; Q9BXX8; Q9BXX9; Q9BXL0; Q9BXL1; Q9BXL2;

AC Q9BXL3; Q9BXL4;

DT 01-NOV-1995 (Rel. 32, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE ATP-binding cassette, sub-family G, member 1 (White protein homolog)  
 DE (ATP-binding cassette transporter 8).  
 GN ABCG1 OR ABC8 OR WHT1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE OF 3-678 FROM N.A. (ISOFORMS 1 AND 4).  
 RC TISSUE=Retina;  
 RX MEDLINE=96256850; PubMed=8659545;  
 RA Chen H.M., Rossier C., Lalioti M.D., Lynn A., Chakravarti A.,  
 RA Perrin G., Antonarakis S.E.;  
 RT "Cloning of the cDNA for a human homologue of the Drosophila white  
 RT gene and mapping to chromosome 21q22.3.";  
 RL Am. J. Hum. Genet. 59:66-75(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=20289799; PubMed=10830953;  
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,  
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,  
 RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,  
 RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,  
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,  
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,  
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,  
 RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,  
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,  
 RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,  
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,  
 RA Lehrach H., Reinhardt R., Yaspo M.-L.;  
 RT "The DNA sequence of human chromosome 21.";  
 RL Nature 405:311-319(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=20408883; PubMed=10950923;  
 RA Berry A., Scott H.S., Kudoh J., Talior I., Korostishevsky M.,  
 RA Wattenhofer M., Guipponi M., Barras C., Rossier C., Shibuya K.,  
 RA Wang J., Kawasaki K., Asakawa S., Minoshima S., Shimizu N.,  
 RA Antonarakis S.E., Bonne-Tamir B.;  
 RT "Refined localization of autosomal recessive nonsyndromic deafness  
 RT DFNB10 locus using 34 novel microsatellite markers, genomic  
 RT structure, and exclusion of six known genes in the region.";  
 RL Genomics 68:22-29(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=21192304; PubMed=11279031;  
 RA Porsch-Oezcuemez M., Langmann T., Heimerl S., Borsukova H.,  
 RA Kaminski W.E., Drobnik W., Honer C., Schumacher C., Schmitz G.;  
 RT "The zinc finger protein 202 (ZNF202) is a transcriptional repressor  
 RT of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene  
 RT expression and a modulator of cellular lipid efflux.";  
 RL J. Biol. Chem. 276:12427-12433(2001).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4; 5; 6 AND 7).  
 RX MEDLINE=21092576; PubMed=11162488;



RA Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,  
 RA Assmann G., Cullen P.;  
 RT "Genomic sequence and structure of the human ABCG1 (ABC8) gene.";  
 RL Biochem. Biophys. Res. Commun. 280:121-131(2001).  
 RN [6]  
 RP SEQUENCE OF 33-678 FROM N.A.  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=97186700; PubMed=9034316;  
 RA Croop J.M., Tiller G.E., Fletcher J.A., Lux M.L., Raab E.,  
 RA Goldenson D., Arciniegas S., Son D., Wu R.;  
 RT "Isolation and characterization of a mammalian homolog of the  
 RT Drosophila white gene.";  
 RL Gene 185:77-85(1997).  
 RN [7]  
 RP INDUCTION, AND PROBABLE FUNCTION.  
 RX MEDLINE=20261604; PubMed=10799558;  
 RA Venkateswaran A., Repa J.J., Lobaccaro J.-M.A., Bronson A.,  
 RA Mangelsdorf D.J., Edwards P.A.;  
 RT "Human white/murine ABC8 mRNA levels are highly induced in  
 RT lipid-loaded macrophages. A transcriptional role for specific  
 RT oxysterols.";  
 RL J. Biol. Chem. 275:14700-14707(2000).  
 RN [8]  
 RP INDUCTION, AND PROBABLE FUNCTION.  
 RX MEDLINE=20105556; PubMed=10639163;  
 RA Klucken J., Buechler C., Orso E., Kaminski W.E.,  
 RA Porsch-Oezcueruemez M., Liebisch G., Kapinsky M., Diederich W.,  
 RA Drobnik W., Dean M., Allikmets R., Schmitz G.;  
 RT "ABCG1 (ABC8), the human homolog of the Drosophila white gene, is a  
 RT regulator of macrophage cholesterol and phospholipid transport.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:817-822(2000).  
 RN [9]  
 RP REVIEW.  
 RX MEDLINE=21474438; PubMed=11590207;  
 RA Schmitz G., Langmann T., Heimerl S.;  
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";  
 RL J. Lipid Res. 42:1513-1520(2001).  
 CC -!- FUNCTION: Transporter involved in macrophage lipid homeostasis. Is  
 CC an active component of the macrophage lipid export complex. Could  
 CC also be involved in intracellular lipid transport processes. The  
 CC role in cellular lipid homeostasis may not be limited to  
 CC macrophages.  
 CC -!- SUBUNIT: May form heterodimers with several heterologous partners  
 CC of the ABCG subfamily.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Predominantly  
 CC localized in the intracellular compartments mainly associated with  
 CC the endoplasmic reticulum (ER) and Golgi membranes.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=7;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=1;  
 CC IsoId=P45844-1; Sequence=Displayed;  
 CC Name=2; Synonyms=J;  
 CC IsoId=P45844-2; Sequence=VSP\_000047, VSP\_000051;  
 CC Name=3; Synonyms=ABDE;  
 CC IsoId=P45844-3; Sequence=VSP\_000048, VSP\_000051;  
 CC Name=4; Synonyms=G;

```

CC      IsoId=P45844-4; Sequence=VSP_000051;
CC      Name=5; Synonyms=F;
CC      IsoId=P45844-5; Sequence=VSP_000049, VSP_000051;
CC      Name=6; Synonyms=HI;
CC      IsoId=P45844-6; Sequence=VSP_000046, VSP_000051;
CC      Name=7; Synonyms=C;
CC      IsoId=P45844-7; Sequence=VSP_000050, VSP_000051;
CC      -!- TISSUE SPECIFICITY: EXPRESSED IN SEVERAL TISSUES.
CC      -!- INDUCTION: Strongly induced in monocyte-derived macrophages during
CC            cholesterol influx. Conversely, mRNA and protein expression are
CC            suppressed by lipid efflux. Induction is mediated by the liver X
CC            receptor/retinoide X receptor (LXR/RXR) pathway.
CC      -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
CC            subfamily.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X91249; CAA62631.1; ALT_INIT.
DR      EMBL; AP001746; BAA95530.1; ALT_INIT.
DR      EMBL; AB038161; BAB13728.2; ALT_INIT.
DR      EMBL; AJ289137; CAC00730.1; ALT_INIT.
DR      EMBL; AJ289138; CAC00730.1; JOINED.
DR      EMBL; AJ289139; CAC00730.1; JOINED.
DR      EMBL; AJ289140; CAC00730.1; JOINED.
DR      EMBL; AJ289141; CAC00730.1; JOINED.
DR      EMBL; AJ289142; CAC00730.1; JOINED.
DR      EMBL; AJ289143; CAC00730.1; JOINED.
DR      EMBL; AJ289144; CAC00730.1; JOINED.
DR      EMBL; AJ289145; CAC00730.1; JOINED.
DR      EMBL; AJ289146; CAC00730.1; JOINED.
DR      EMBL; AJ289147; CAC00730.1; JOINED.
DR      EMBL; AJ289148; CAC00730.1; JOINED.
DR      EMBL; AJ289149; CAC00730.1; JOINED.
DR      EMBL; AJ289150; CAC00730.1; JOINED.
DR      EMBL; AJ289151; CAC00730.1; JOINED.
DR      EMBL; AF323658; AAK28836.1; -.
DR      EMBL; AF323644; AAK28836.1; JOINED.
DR      EMBL; AF323645; AAK28836.1; JOINED.
DR      EMBL; AF323646; AAK28836.1; JOINED.
DR      EMBL; AF323647; AAK28836.1; JOINED.
DR      EMBL; AF323648; AAK28836.1; JOINED.
DR      EMBL; AF323649; AAK28836.1; JOINED.
DR      EMBL; AF323650; AAK28836.1; JOINED.
DR      EMBL; AF323651; AAK28836.1; JOINED.
DR      EMBL; AF323652; AAK28836.1; JOINED.
DR      EMBL; AF323653; AAK28836.1; JOINED.
DR      EMBL; AF323654; AAK28836.1; JOINED.
DR      EMBL; AF323655; AAK28836.1; JOINED.
DR      EMBL; AF323656; AAK28836.1; JOINED.
DR      EMBL; AF323657; AAK28836.1; JOINED.
DR      EMBL; AF323664; AAK28842.1; -.

```

DR EMBL; AF323658; AAK28833.1; -.  
 DR EMBL; AF323640; AAK28833.1; JOINED.  
 DR EMBL; AF323645; AAK28833.1; JOINED.  
 DR EMBL; AF323646; AAK28833.1; JOINED.  
 DR EMBL; AF323647; AAK28833.1; JOINED.  
 DR EMBL; AF323648; AAK28833.1; JOINED.  
 DR EMBL; AF323649; AAK28833.1; JOINED.  
 DR EMBL; AF323650; AAK28833.1; JOINED.  
 DR EMBL; AF323651; AAK28833.1; JOINED.  
 DR EMBL; AF323652; AAK28833.1; JOINED.  
 DR EMBL; AF323653; AAK28833.1; JOINED.  
 DR EMBL; AF323654; AAK28833.1; JOINED.  
 DR EMBL; AF323655; AAK28833.1; JOINED.  
 DR EMBL; AF323656; AAK28833.1; JOINED.  
 DR EMBL; AF323657; AAK28833.1; JOINED.  
 DR EMBL; AF323660; AAK28838.1; -.  
 DR EMBL; AF323663; AAK28841.1; ALT\_INIT.  
 DR EMBL; AF323658; AAK28835.1; -.  
 DR EMBL; AF323642; AAK28835.1; JOINED.  
 DR EMBL; AF323645; AAK28835.1; JOINED.  
 DR EMBL; AF323646; AAK28835.1; JOINED.  
 DR EMBL; AF323647; AAK28835.1; JOINED.  
 DR EMBL; AF323648; AAK28835.1; JOINED.  
 DR EMBL; AF323649; AAK28835.1; JOINED.

Query Match 17.4%; Score 586; DB 1; Length 678;  
 Best Local Similarity 28.3%; Pred. No. 1.3e-34;  
 Matches 165; Conservative 125; Mismatches 240; Indels 54; Gaps 15;

Qy 23 SSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKSCQQKWDRQILKDVSLYIESGQI 82  
 ||| : : | | : ||| ||| : : | : | || : :  
 Db 67 SSLPRRAAVNIEFR-----DLSYSVPE--GPWW-----RKKGYKTLKKGISGKFNSGEL 113  
  
 Qy 83 MCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVCCELRRDQ--FQDCFSYVLQSDVFL 140  
 : | : | || : || : : : | | | : | | : | | : | : | : | : |  
 Db 114 VAIMGPSGAGKSTLMNILAG-YRETG-MKGAVLING--LPRDLRCFRKVSICYIMQDDMLL 169  
  
 Qy 141 SSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTELSLSHVADQMIGSYNFGGISSGERR 200  
 ||| : : | | : : : : : | | : | : | : | : | : | : | :  
 Db 170 PHLTVQEAMMVSAHLKL-QEKDEGRREMVKELTALGLLSCANTRTGS-----LSGGQRK 223  
  
 Qy 201 RVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQH 260  
 | : : | : : | | | | : | : | : | : | : | : | : | : | : | :  
 Db 224 RLAI ALELVNNPPVMFFDEPTSGLDASCFQVVSIMKGLAQGGRSIICTIHQPSAKLFEL 283  
  
 Qy 261 FDKIAILTYGELVFCGTPEEMLGFFNNGCYPCPEHSNPFDFYMDLTSVDTQSREREIETY 320  
 || : : | : | : | : : : | | : | | | : : | : : : :  
 Db 284 FDQLYVLSQGQCVRGKVCNLPYLRDLGLNCPTYHNPADFVMEVASGEYGDQNSRLVRA 343  
  
 Qy 321 KRVQMLECAFKES-----DIYHKILENIERARYLKTLPMPFKTKDPPGMFG----- 367  
 | | : | : : : | : : : | | | | | | | | | | | | | | | | |  
 Db 344 VREGMCDSDHKRDLGGDAEVNPFLLWHRPSEEVKQTKRLKGL-----RKDSSSMEGCHSF 397  
  
 Qy 368 -----KLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNTLKGAVQDRV 421  
 : : | : | : : | : : : : | : | : | | | | | | | | | |  
 Db 398 SASCLTQFCILFKRTFLSIMRDSVLTHLRITSHIGIGLLIGLLYLGIGNEAK--VLSNS 455

Qy 422 GLLYQLVGATPYTGMLNAVNLFPMRLRAVSDQESQDGLYHKWQMLLAYVLHVLPPFSVIATV 481  
 | |: : : :: | ||: | :| : | || : :|| :: |  
 Db 456 GFLFFSMLFLMFAALMPTVLTFFLEMGVFLREHLNYWYSLKAYYLAKTMADVFPQIMFPV 515  
 Qy 482 IFSSVCYWTLGLYPEVARFGYFSAALLAPHLIGEFLLTVLLGIVQNPNIIVNSIVALLSIS 541  
 : |: || : || |:| |: : | | |:| | : | ::  
 Db 516 AYCSIVYWMTSQPSDAVRVFLFAALGTMTSLVAQSLGL-LIGAASLQVATFVGVPVTAI 574  
 Qy 542 GLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLN 585  
 :|: ||| : :| |: : | :: :| | :::: |||:  
 Db 575 PVLLFSGFFVSFDTIPTYLQWMSYISYVRYGFEGVILS-IYGLD 617

# RESULT 12

YPC3\_CAEEL

ID YPC3\_CAEEL STANDARD; PRT; 598 AA.

AC Q11180;

DT 01-NOV-1997 (Rel. 35, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Putative ABC transporter C05D10.3 in chromosome III.

GN C05D10.3.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Du Z.;

RL Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.

RN [2]

RP REVISIONS.

RA Waterston R.;

RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; U13645; AAA20989.2; -.  
 DR WormPep; C05D10.3; CE29170.

DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR005284; Pigment\_permease.  
 DR Pfam; PF00005; ABC\_tran; 1.

DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.

DR TIGRFAMs; TIGR00955; 3a01204; 1.

DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; FALSE\_NEG.

DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.

KW Hypothetical protein; ATP-binding; Transmembrane; Transport.  
 FT NP\_BIND 27 34 ATP (POTENTIAL).  
 FT TRANSMEM 336 356 POTENTIAL.  
 FT TRANSMEM 425 445 POTENTIAL.  
 FT TRANSMEM 453 473 POTENTIAL.  
 FT TRANSMEM 478 498 POTENTIAL.  
 SQ SEQUENCE 598 AA; 66906 MW; 9D6414E06898E343 CRC64;

Query Match 17.3%; Score 584.5; DB 1; Length 598;  
 Best Local Similarity 25.4%; Pred. No. 1.4e-34;  
 Matches 154; Conservative 137; Mismatches 273; Indels 43; Gaps 12;

Qy 67 RQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVCCELRRDQF 126  
 :|| :|| ||::: |||||:||||: :| :| : :| : :  
 Db 7 KEILHNVSMAESGKLLAILGSSGAGKTTLMNVLTSRNLTNLDVQGSILIDGRRANKWKI 66

Qy 127 QDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNK-----KVEAVMTLSLSHV 181  
 : :| | |: :| || |: | | :| :| | |: :|  
 Db 67 REMSAFVQQHDMFVGTMAREHLQFMARL---RMGDQYYSDHERQLRVEQVLTQMGLKKC 123

Qy 182 ADQMIGSYN-FGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240  
 || :|| | |:| ||:::| |:| | |:| |||:| | :| | ||  
 Db 124 ADTVIGIPNQLKGLSCGEKKRLSFASEILTCPKILFCDEPTSGLD AFMAGHV VQALRSLA 183

Qy 241 RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNGCYPCPEHSNPF 300  
 ||:|||| | : : : : | : : | : : | ||||| : |||  
 Db 184 DNGMTVIIITIHQPSSHVYSLFNNVCLMACGRVIYLGPGDQAVPLFEKCGYPCPAYNPAD 243

Qy 301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPV----- 355  
 : :| | :| : : | :| :| | | :| |  
 Db 244 HLIRTLAVIDSDRATSMKTISKIRQ---GFLSTD LGQSVLA-IGNANKLRAASFVTGSDT 299

Qy 356 -----PFKTKDPPGMF-GKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLF--LIFYLLR 407  
 | :| | : | | :| : :||:| || :| :  
 Db 300 SEKTKTFFNQDYNASFWTQFLALFWRSWLTVIRDPNLLSVRLQLILITAFITGIVFF--- 356

Qy 408 VQNNTLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPMRAVSDQ-----ESQDGLYHK 461  
 | : | : : : :| : : || : : : | :| :|  
 Db 357 -QTPVTPATIISINGIMFN-----HIRNMNFMQLQFPNVPVITAELPIVLRENANGVYRT 409

Qy 462 WQMLLAYVLHVLFPFSVIATVIFSSVCYWTGLYPEVAREFGYFSAALLAPHLIGEFLLVL 521  
 || : || :| : : : || ||| : : | : : : :  
 Db 410 SAYFLAKNIAELPQYIILPILYNTIVYWMGLYPNFWNYCFASLVITILITNVAISISYAV 469

Qy 522 LGIVQNPNIIVNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVNEF 581  
 | | : :| : : : : | || :| | | : : || | | :| :  
 Db 470 ATIFANTDVAMTILPIFVVPIMAFG-GFFITFDAIPSYFKWLSSLSYFKYGYEALAIN 528

Qy 582 YGLNFTCGGSNTSMLNHPMAITQGVQFIEKTCPGATS RFTANFLILYGFIPALVILGIV 641  
 : | :|| : : : : : : | : ||:| :| :|  
 Db 529 DSIKVIPECFNSSMTAFALDSCPKNGHQVLESIDFSASHKIFDISILFGMFIGIRIIAYV 588

Qy 642 IFKVRDY 648  
 :| |  
 Db 589 ALLIRSY 595

## RESULT 13

WHIT\_LUCCU

ID WHIT\_LUCCU STANDARD; PRT; 677 AA.  
AC Q05360;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE White protein.  
GN W.  
OS *Lucilia cuprina* (Greenbottle fly) (Australian sheep blowfly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;  
OC Calliphoridae; *Lucilia*.  
OX NCBI\_TaxID=7375;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97087158; PubMed=8933176;  
RA Garcia R.L., Perkins H.D., Howells A.J.;  
RT "The structure, sequence and developmental pattern of expression of  
RT the white gene in the blowfly *Lucilia cuprina*."  
RL Insect Mol. Biol. 5:251-260(1996).  
RN [2]  
RP SEQUENCE OF 490-584 FROM N.A.  
RX MEDLINE=90264941; PubMed=1971656;  
RA Elizur A., Vacek A.T., Howells A.J.;  
RT "Cloning and characterization of the white and topaz eye color genes  
RT from the sheep blowfly *Lucilia cuprina*."  
RL J. Mol. Evol. 30:347-358(1990).  
CC -!- FUNCTION: May be part of a membrane-spanning permease system  
CC necessary for the transport of pigment precursors into pigment  
CC cells responsible for eye color.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U38899; AAA82057.1; -.  
DR EMBL; X53265; CAA37365.1; -.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003439; ABC\_transporter.  
DR InterPro; IPR005284; Pigment\_permease.  
DR Pfam; PF00005; ABC\_tran; 1.  
DR ProDom; PD000006; ABC\_transporter; 1.  
DR SMART; SM00382; AAA; 1.  
DR TIGRFAMs; TIGR00955; 3a01204; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
KW Pigment; ATP-binding; Transmembrane; Transport.  
FT NP\_BIND 119 126 ATP (POTENTIAL).  
FT TRANSMEM 431 451 POTENTIAL.  
FT TRANSMEM 456 476 POTENTIAL.

FT TRANSMEM 506 526 POTENTIAL.  
 FT TRANSMEM 534 554 POTENTIAL.  
 FT TRANSMEM 563 583 POTENTIAL.  
 FT TRANSMEM 647 667 POTENTIAL.  
 SQ SEQUENCE 677 AA; 75365 MW; D16FC11C97EED51D CRC64;

Query Match 17.3%; Score 583.5; DB 1; Length 677;  
 Best Local Similarity 28.6%; Pred. No. 2e-34;  
 Matches 191; Conservative 132; Mismatches 267; Indels 79; Gaps 21;

Qy 20 GSLSSLEQGSVTGTEARHSL---GVLHVSYS----VSNRV-GPWWNIKSCQQKWDRQILK 71  
 ||| | | | : : | : | : ||| | : | : | : : |  
 Db 45 GSLVSNESASEKLTYSWCNLDVFGEVHQPGSNWKQLVNRVKGVFCNERHI-PKPRKHLIK 103

Qy 72 DVSLYIESGQIMCILGSSSGSKTTLLDAISGRLLRRTGTLEGEV--FVNGCELRRDQFQDC 129  
 :| |:: :|||:|||||:|:: | | : :|| : : |  
 Db 104 NVCGVAYPGELLAVMGSSGAGKTTLLNALAFRSARGVQISPSSVRMLNGHPVDAKEMQAR 163

Qy 130 FSYVLQSDVFLSSLTVRETLYRTAMLALCRSSADFYN-KKVEAVMTELSLSHVADQMIG- 187  
 :|| | |:|: ||| || | : | : : | : :|: |: :||| : :||  
 Db 164 CAYVQDDDLFIGSLTAREHLIFQATVRMPRTMTQKQKLQRVDQVIQDLSLIKQNTIIGV 223

Qy 188 SYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLLAELARRDRIVI 247  
 |:| |||:|:: |:| | || :: |||:| || | :| :| :|::| : ||  
 Db 224 PGRVKGLSGGERKRLAFASEALTDPELLICDEPTSGLDSEMAASVVQVLKKLSQRGKTVI 283

Qy 248 VTIHQPRSELFQHFDKIAILTYGELVFCGTPEEMLGFFNNGGYPCPEHSNPFDFYMDLTS 307  
 :|||| | |||: | ||| :: | : | || | : ||: | || : || |||: : :  
 Db 284 LTIHQPSSELFELFDKILLMAEGRVAF LGTPVEAVDFFSFIGAQCPNTYNPADFYVQVLA 343

Qy 308 VDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPVFKT----KD-- 361  
 | | |||: |: : | : : :| : : : || ||  
 Db 344 V---VPGREIESRDRISKICDNFAVGKVSREMEQNFQK-----IAAKTDGLQKDDE 391

Qy 362 -----PPGMFGKLGVLRLRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNLTKGA 416  
 | : : : | : : : | :||:| : : : || : | | |  
 Db 392 TTILYKASWFTQFRAIMWRSWISTLKEPLLVKVRLIQTMVAV-LIGLIFLNQPMTQVG- 449

Qy 417 VQDRVGLLYQLVGATPYTGMLNAVNLFPMRLRAVSDQESQDGLYHKWQMLLAYVLHVLPPFS 476  
 | : | : : : : : :|:| | :||: || | | ||  
 Db 450 VMNINGAIFLFLTNMTFQNVFAVINVFTSELPVFMRETRSRRLYRCDTYFLGKTLAELPLF 509

Qy 477 VIATVIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGEFLT---LVLLGIVQNPNI VNS 533  
 :: :| : : | :|| | : :|:| | : : : : : : : : |  
 Db 510 LVVPFLFIAIAYPMIGLRPGIT---HFLSALALVTLVANVSTSFGYLISCSTSTSMALS 566

Qy 534 IVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGL---NFTCGG 590  
 : | :| || | | : | :|: | | ||: :| | |::| : : ||  
 Db 567 VGPPLTIPFLLEGGVFL-NSGSVPVYFKWLSYFSWFRYANEGLLINQWADVQPGEITCTS 625

Qy 591 SNTSMLNHPMCAITQGVQFIEKTCF--GATSRFTANF-----LILYGFI PALVILGIVIF 643  
 :|| | || | | || | || :|| | :||  
 Db 626 TNT-----TCPSSGXVXLETNLNFRDKFTFRLYG-----LILLILIF 661

Qy 644 KVRDY LISR 652  
 :: | : :  
 Db 662 RIAGYVAXK 670

# RESULT 14

## WHIT\_ANOGA

ID WHIT\_ANOGA STANDARD; PRT; 695 AA.  
AC Q27256; Q17006;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE White protein.  
GN W.  
OS Anopheles gambiae (African malaria mosquito).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.  
OX NCBI\_TaxID=7165;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Suakoko / G3;  
RX MEDLINE=96423158; PubMed=8825759;  
RA Besansky N.J., Bedell J.A., Benedict M.Q., Mukabayire O., Hilfiker D.,  
RA Collins F.H.;  
RT "Cloning and characterization of the white gene from Anopheles  
RT gambiae.";  
RL Insect Mol. Biol. 4:217-231(1995).  
CC -!- FUNCTION: May be part of a membrane-spanning permease system  
CC necessary for the transport of pigment precursors into pigment  
CC cells responsible for eye color.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U29486; AAC46995.1; -.  
DR EMBL; U29485; AAC46994.1; -.  
DR EMBL; U29484; AAC47423.1; -.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003439; ABC\_transporter.  
DR InterPro; IPR008965; Cellul\_bind.  
DR InterPro; IPR005284; Pigment\_permease.  
DR Pfam; PF000005; ABC\_tran; 1.  
DR ProDom; PD000006; ABC\_transporter; 1.  
DR SMART; SM00382; AAA; 1.  
DR TIGRFAMs; TIGR00955; 3a01204; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
KW Pigment; ATP-binding; Transmembrane; Transport.  
FT NP\_BIND 133 140 ATP (POTENTIAL).  
FT NP\_BIND 288 295 ATP (POTENTIAL).  
FT TRANSMEM 444 464 POTENTIAL.  
FT TRANSMEM 474 494 POTENTIAL.  
FT TRANSMEM 524 544 POTENTIAL.



FT	TRANSMEM	552	572	POTENTIAL.
FT	TRANSMEM	581	601	POTENTIAL.
FT	TRANSMEM	669	689	POTENTIAL.
FT	CARBOHYD	472	472	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	645	645	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	100	100	N -> S (IN REF. 1; AAC47423).
FT	CONFLICT	691	693	SRS -> YAR (IN REF. 1; AAC47423).
SO	SEQUENCE	695 AA;	77218 MW;	EE8B9517239B2961 CRC64;

Qy	158	IKSC--QQKWD----RQILKDVSLYIESGQIMCILGSSSGSGKTTLLDAISGRRLRTGTLE	111
Db	98	LRNCCTRQRKDFNPRKHLLKNVTGVAKSGELLAVMGSSGAGKTTLLNALAFRSPPGVKIS	157
Qy	112	GEVF--VNGCELRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRS-SADFYNNK	168
Db	158	PNAVRALNGVPVNAEQLRARCAIVQQDDLFIPSLTTREHLLFQAMLRMGRDVPASVKQHR	217
Qy	169	VEAVMTELSLSHVADQMIGS-YNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCM	227
Db	218	VQEVQLQELSLVKCADTIIGAPGRIKGLSGGERKRLAFASETLTDPHLLLCDEPTSGLDSE	277
Qy	228	TANQIVLLLAELARRDRIVIVTIHQPRSELFQHFQDKIALTYGELVFCGTPEEMLGFFNN	287
Db	278	MAHSVQLQVLKGMAMKGKTIILTIHQPSSELYCLFDKILLVAEGRVAFLGSPYQSAEFFSQ	337
Qy	288	CGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERAR	347
Db	338	LGIPCPPNYNPADFYVQMLAI---APAKEAECDRMIKKICDSFAVSPiAREVLETAS---	391
Qy	348	YLKTLPMVPFKTKDPPGMF-----GKLG-----VLLRRVTRNLMRNKQAVIMRL	391
Db	392	-----VAGKGMDEPYMLQQVEGVGSTGYRSSWWTQFYCILWRWSVLKDPMLVKVRL	444
Qy	392	VQNLIMGLFLIFYLLRVQNNTLKGAV-----QDRV-----GLLYQLVGATPYTGMLNAV	440
Db	445	LQTAMVA-----TLIGSIYFGQVLDQDGMVNINGSLLFLFLTNMTFQNVFAVI	491
Qy	441	NLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLPPFSVIATVIFSSVCYWTGLYPEVARF	500
Db	492	NVFSaelPVFLREKRSRLYRVDTYFLGKTIAELPLFIAPVFFVTSITYPMIGL----RT	546
Qy	501	GYFSAALLAPHLIGEFLLTVLLGIVQNP-----IVNSIVALLSIS-----GLLIG	546
Db	547	G-----ATHYL---TTLFIIVTLVANVSTSFGYLISCASSISMALSVGPPVVIPLIF	596
Qy	547	SGFIRNIQEMPIPLKILGYFTFQKYCCeilVVNEFYGL---NFTCGGSNTSMLNHPMCAI	603
Db	597	GGFFLNSASVPAYFKYLSYLSWFRYANEALLINQWSTVVDGEIACTRANV-----	646
Qy	604	TQGVQFIEKTCPGATSrFTANFLILYGFIPALVILGIVIFKVRDYLI	650
Db	647	-----TCPRSE-----IILETFNFRVEDFAL	667

RESULT 15

ABG4 HUMAN

ID ABG4 HUMAN STANDARD; PRT; 646 AA.  
AC Q9H172;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE ATP-binding cassette, sub-family G, member 4.  
GN ABCG4 OR WHITE2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21518231; PubMed=11606068;  
RA Engel T., Lorkowski S., Lueken A., Rust S., Schlueter B., Berger G.,  
RA Cullen P., Assmann G.;  
RT "The human ABCG4 gene is regulated by oxysterols and retinoids in  
RT monocyte-derived macrophages.";  
RL Biochem. Biophys. Res. Commun. 288:483-488(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RP SEQUENCE OF 20-646 FROM N.A.  
RC TISSUE=Dorsal root ganglion;  
RX MEDLINE=22170423; PubMed=12183068;  
RA Oldfield S., Lowry C., Ruddick J., Lightman S.;  
RT "ABCG4: a novel human white family ABC-transporter expressed in the  
RT brain and eye.";  
RL Biochim. Biophys. Acta 1591:175-179(2002).  
CC -!- FUNCTION: May be involved in macrophage lipid homeostasis.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)

```

CC      subfamily.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ308237; CAC87131.1; -.
DR EMBL; BC041091; AAH41091.1; -.
DR EMBL; AJ300465; CAC17140.1; -.
DR PIR; JC7777; JC7777.
DR Genew; HGNC:13884; ABCG4.
DR MIM; 607784; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Glycoprotein; Transmembrane; Transport.
FT DOMAIN      1      393      CYTOPLASMIC (POTENTIAL).
FT TRANSMEM     394     414      1 (POTENTIAL).
FT DOMAIN      415     425      EXTRACELLULAR (POTENTIAL).
FT TRANSMEM     426     446      2 (POTENTIAL).
FT DOMAIN      447     472      CYTOPLASMIC (POTENTIAL).
FT TRANSMEM     473     493      3 (POTENTIAL).
FT DOMAIN      494     503      EXTRACELLULAR (POTENTIAL).
FT TRANSMEM     504     524      4 (POTENTIAL).
FT DOMAIN      525     532      CYTOPLASMIC (POTENTIAL).
FT TRANSMEM     533     553      5 (POTENTIAL).
FT DOMAIN      554     617      EXTRACELLULAR (POTENTIAL).
FT TRANSMEM     618     638      6 (POTENTIAL).
FT DOMAIN      639     646      CYTOPLASMIC (POTENTIAL).
FT NP_BIND      102     109      ATP (POTENTIAL).
FT CARBOHYD     422     422      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE     646 AA;  71895 MW;  9CCEC6E150772611 CRC64;

```

```

Query Match          16.9%;  Score 569.5;  DB 1;  Length 646;
Best Local Similarity 26.0%;  Pred. No. 1.9e-33;
Matches 174;  Conservative 138;  Mismatches 273;  Indels 83;  Gaps 19;

```

```

Qy      10 EGARGP----HINRGSLSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKSCQQKW 65
      :|| |   |: :   |   :   |::: | || || |   :::
Db      26 DGAEPVLTTHLKKVENHITEAQRFSHLPKRSAVDIEFVELSYSVREGPCW-----RKRG 80

Qy      66 DRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVVFVNG--CELRR 123
      : :|| :|   ::: |:| ||::|:| :: :| | :|:: ||| |||
Db      81 YKTLKCLSGKFCRRELIGIMGPSGAGKSTFMNILAG-YRESG-MKGQILVNGRPRELRT 138

Qy      124 DQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTESL---SH 180
      : | |:| |:| ||| | : :| |   : :| :| | | ||
Db      139 FRKMSC--YIMQDDMLLPHLTVLEAMVSVANLKLSEKQ-EVKKELVTEILTALGLMSCSH 195

```

